DEX-0273 CE

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 September 2002 (12.09.2002)

PCT

(10) International Publication Number
WO 02/070709 A2

(51) International Patent Classification[7]: C12N 15/12, 5/10, C12Q 1/68, C07K 14/47, 16/18, G01N 33/50, 33/68, A01K 67/027, A61K 38/17, 39/395

(21) International Application Number: PCT/US02/03709

(22) International Filing Date: 8 February 2002 (08.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/268,117	9 February 2001 (09.02.2001)	US
60/269,618	15 February 2001 (15.02.2001)	US
60/271,118	23 February 2001 (23.02.2001)	US
60/274,486	7 March 2001 (07.03.2001)	US
60/274,436	9 March 2001 (09.03.2001)	US
60/334,229	28 November 2001 (28.11.2001)	US
60/353,284	1 February 2002 (01.02.2002)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandero, CA 94577 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 40633 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street, #205, San Leandro, CA 94577 (US). **ELLIOT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **HONCHELL, Cynthia, D.** [US/US]; 400 Laurel Street, #203, San Carlos, CA 94070 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **DING, Li** [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94506 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **GHANDI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue, #23, Mountain View, CA 94043 (US). **SANJANWALA, Madhusudan, M.** [US/US]; 210 Silvia Court, Los Altos, CA 94024 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **SWARNAKER, Anita** [CA/US]; 8 Locksley Avenue, #5D, San Francisco, CA 94122 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **SAPPERSTEIN, Stephanie, K.** [US/US]; 3531 Highland Avenue, Redwood City, CA 94062 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **MARQUIS, Joseph, P.** [US/US]; 3299 Moulin Lane, San Jose, CA 95135 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: MOLECULES FOR DISEASE DETECTION AND TREATMENT

(57) Abstract: The invention provides full-length human molecules for disease detection and treatment (MDDT) and polynucleotides which identify and encode MDDT. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of MDDT.

WO 02/070709 A2

MOLECULES FOR DISEASE DETECTION AND TREATMENT

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of full-length human molecules for disease detection and treatment and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, autoimmune/inflammatory, developmental, neurological, and cardiovascular disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of full-length human molecules for disease detection and treatment.

BACKGROUND OF THE INVENTION

It is estimated that only 2% of mammalian DNA encodes proteins, and only a small fraction of the genes that encode proteins is actually expressed in a particular cell at any time. The various types of cells in a multicellular organism differ dramatically both in structure and function, and the identity of a particular cell is conferred by its unique pattern of gene expression. In addition, different cell types express overlapping but distinctive sets of genes throughout development. Cell growth and proliferation, cell differentiation, the immune response, apoptosis, and other processes that contribute to organismal development and survival are governed by regulation of gene expression. Appropriate gene regulation also ensures that cells function efficiently by expressing only those genes whose functions are required at a given time. Factors that influence gene expression include extracellular signals that mediate cell-cell communication and coordinate the activities of different cell types. Gene expression is regulated at the level of DNA and RNA transcription, and at the level of mRNA translation.

Aberrant expression or mutations in genes and their products may cause, or increase susceptibility to, a variety of human diseases such as cancer and other cell proliferative disorders. The identification of these genes and their products is the basis of an ever-expanding effort to find markers for early detection of diseases and targets for their prevention and treatment. For example, cancer represents a type of cell proliferative disorder that affects nearly every tissue in the body. The development of cancer, or oncogenesis, is often correlated with the conversion of a normal gene into a cancer-causing gene, or oncogene, through abnormal expression or mutation. Oncoproteins, the products of oncogenes, include a variety of molecules that influence cell proliferation, such as growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. In contrast, tumor-suppressor genes are involved in inhibiting cell

proliferation. Mutations which reduce or abrogate the function of tumor-suppressor genes result in aberrant cell proliferation and cancer. Thus a wide variety of genes and their products have been found that are associated with cell proliferative disorders such as cancer, but many more may exist that are yet to be discovered.

DNA-based arrays can provide an efficient, high-throughput method to examine gene expression and genetic variability. For example, SNPs, or single nucleotide polymorphisms, are the most common type of human genetic variation. DNA-based arrays can dramatically accelerate the discovery of SNPs in hundreds and even thousands of genes. Likewise, such arrays can be used for SNP genotyping in which DNA samples from individuals or populations are assayed for the presence of selected SNPs. These approaches will ultimately lead to the systematic identification of all genetic variations in the human genome and the correlation of certain genetic variations with disease susceptibility, responsiveness to drug treatments, and other medically relevant information. (See, for example, Wang, D.G. et al. (1998) Science 280:1077-1082.)

DNA-based array technology is especially important for the rapid analysis of global gene expression patterns. For example, genetic predisposition, disease, or therapeutic treatment may directly or indirectly affect the expression of a large number of genes in a given tissue. In this case, it is useful to develop a profile, or transcript image, of all the genes that are expressed and the levels at which they are expressed in that particular tissue. A profile generated from an individual or population affected with a certain disease or undergoing a particular therapy may be compared with a profile generated from a control individual or population. Such analysis does not require knowledge of gene function, as the expression profiles can be subjected to mathematical analyses which simply treat each gene as a marker. Furthermore, gene expression profiles may help dissect biological pathways by identifying all the genes expressed, for example, at a certain developmental stage, in a particular tissue, or in response to disease or treatment. (See, for example, Lander, E.S. et al. (1996) Science 274:536-539.)

Certain genes are known to be associated with diseases because of their chromosomal location, such as the genes in the myotonic dystrophy (DM) regions of mouse and human. The mutation underlying DM has been localized to a gene encoding the DM-kinase protein, but another active gene, DMR-N9, is in close proximity to the DM-kinase gene (Jansen, G. et al. (1992) Nat. Genet. 1:261-266). DMR-N9 encodes a 650 amino acid protein that contains WD repeats, motifs found in cell signaling proteins. DMR-N9 is expressed in all neural tissues and in the testis, suggesting a role for DMR-N9 in the manifestation of mental and testicular symptoms in severe cases of DM (Jansen, G. et al. (1995) Hum. Mol. Genet. 4:843-852).

Other genes are identified based upon their expression patterns or association with disease syndromes. For example, autoantibodies to subcellular organelles are found in patients with systemic rheumatic diseases. A recently identified protein, golgin-67, belongs to a family of Golgi autoantigens having alpha-helical coiled-coil domains (Eystathioy, T. et al. (2000) J. Autoimmun. 14:179-187). The Stac gene was identified as a brain specific, developmentally regulated gene. The Stac protein contains an SH3 domain, and is thought to be involved in neuron-specific signal transduction (Suzuki, H. et al. (1996) Biochem. Biophys. Res. Commun. 229:902-909).

Calponin is an actin-binding protein that may participate in the function and organization the cytoskeleton (Takahashi, K. et al. (1986) Biochem. Biophys. Res. Commun. 141:20-26). The N-terminus of calponin can interact with calcium-binding proteins and tropomyosin. Also at located at the N-terminus is the CH-domain (calponin homology domain) that is found within the structure of several additional actin-binding proteins (Gusev, N.B. (2001) Biochemistry (Mosc) 66:1112-1121).

Secreted Proteins

Protein transport and secretion are essential for cellular function. Protein transport is mediated by a signal peptide located at the amino terminus of the protein to be transported or secreted. The signal peptide is comprised of about ten to twenty hydrophobic amino acids which target the nascent protein from the ribosome to a particular membrane bound compartment such as the endoplasmic reticulum (ER). Proteins targeted to the ER may either proceed through the secretory pathway or remain in any of the secretory organelles such as the ER, Golgi apparatus, or lysosomes. Proteins that transit through the secretory pathway are either secreted into the extracellular space or retained in the plasma membrane. Proteins that are retained in the plasma membrane contain one or more transmembrane domains, each comprised of about 20 hydrophobic amino acid residues. Secreted proteins are generally synthesized as inactive precursors that are activated by post-translational processing events during transit through the secretory pathway. Such events include glycosylation, proteolysis, and removal of the signal peptide by a signal peptidase. Other events that may occur during protein transport include chaperone-dependent unfolding and folding of the nascent protein and interaction of the protein with a receptor or pore complex. Examples of secreted proteins with amino terminal signal peptides are discussed below and include proteins with important roles in cell-to-cell signaling. Such proteins include transmembrane receptors and cell surface markers, extracellular matrix molecules, cytokines, hormones, growth and differentiation factors, enzymes, neuropeptides, vasomediators, cell surface markers, and antigen recognition molecules. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of The Cell, Garland Publishing, New York, NY, pp. 557-560, 582-592.)

Cell surface markers include cell surface antigens identified on leukocytic cells of the immune system. These antigens have been identified using systematic, monoclonal antibody (mAb)-based "shot gun" techniques. These techniques have resulted in the production of hundreds of mAbs directed against unknown cell surface leukocytic antigens. These antigens have been grouped into "clusters of differentiation" based on common immunocytochemical localization patterns in various differentiated and undifferentiated leukocytic cell types. Antigens in a given cluster are presumed to identify a single cell surface protein and are assigned a "cluster of differentiation" or "CD" designation. Some of the genes encoding proteins identified by CD antigens have been cloned and verified by standard molecular biology techniques. CD antigens have been characterized as both transmembrane proteins and cell surface proteins anchored to the plasma membrane via covalent attachment to fatty acid-containing glycolipids such as glycosylphosphatidylinositol (GPI). (Reviewed in Barclay, A.N. et al. (1995) The Leucocyte Antigen Facts Book, Academic Press, San Diego, CA, pp. 17-20.)

Matrix proteins (MPs) are transmembrane and extracellular proteins which function in formation, growth, remodeling, and maintenance of tissues and as important mediators and regulators of the inflammatory response. The expression and balance of MPs may be perturbed by biochemical changes that result from congenital, epigenetic, or infectious diseases. In addition, MPs affect leukocyte migration, proliferation, differentiation, and activation in the immune response. MPs are frequently characterized by the presence of one or more domains which may include collagen-like domains, EGF-like domains, immunoglobulin-like domains, and fibronectin-like domains. In addition, MPs may be heavily glycosylated and may contain an Arginine-Glycine-Aspartate (RGD) tripeptide motif which may play a role in adhesive interactions. MPs include extracellular proteins such as fibronectin, collagen, galectin, vitronectin and its proteolytic derivative somatomedin B; and cell adhesion receptors such as cell adhesion molecules (CAMs), cadherins, and integrins. (Reviewed in Ayad, S. et al. (1994) The Extracellular Matrix Facts Book, Academic Press, San Diego, CA, pp. 2-16; Ruoslahti, E. (1997) Kidney Int. 51:1413-1417; Sjaastad, M.D. and Nelson, W.J. (1997) BioEssays 19:47-55.)

Mucins are highly glycosylated glycoproteins that are the major structural component of the mucus gel. The physiological functions of mucins are cytoprotection, mechanical protection, maintenance of viscosity in secretions, and cellular recognition. MUC6 is a human gastric mucin that is also found in gall bladder, pancreas, seminal vesicles, and female reproductive tract (Toribara, N.W. et al. (1997) J. Biol. Chem. 272:16398-16403). The MUC6 gene has been mapped to human chromosome 11 (Toribara, N.W. et al. (1993) J. Biol. Chem. 268:5879-5885). Hemomucin is a novel

Drosophila surface mucin that may be involved in the induction of antibacterial effector molecules (Theopold, U. et al. (1996) J. Biol. Chem. 217:12708-12715).

Tuftelins are one of four different enamel matrix proteins that have been identified so far. The other three known enamel matrix proteins are the amelogenins, enamelin and ameloblastin. Assembly of the enamel extracellular matrix from these component proteins is believed to be critical in producing a matrix competent to undergo mineral replacement. (Paine, C.T. et al. (1998) Connect Tissue Res. 38:257-267). Tuftelin mRNA has been found to be expressed in human ameloblastoma tumor, a non-mineralized odontogenic tumor (Deutsch, D. et al. (1998) Connect. Tissue Res. 39:177-184).

Olfactomedin-related proteins are extracellular matrix, secreted glycoproteins with conserved C-terminal motifs. They are expressed in a wide variety of tissues and in broad range of species, from *Caenorhabditis elegans* to *Homo sapiens*. Olfactomedin-related proteins comprise a gene family with at least 5 family members in humans. One of the five, TIGR/myocilin protein, is expressed in the eye and is associated with the pathogenesis of glaucoma (Kulkarni, N.H. et al. (2000) Genet. Res. 76:41-50). Research by Yokoyama et al. (1996) found a 135-amino acid protein, termed AMY, having 96% sequence identity with rat neuronal olfactomedin-releated ER localized protein in a neuroblastoma cell line cDNA library, suggesting an essential role for AMY in nerve tissue (Yokoyama, M. et al. (1996) DNA Res. 3:311-320). Neuron-specific olfactomedin-related glycoproteins isolated from rat brain cDNA libraries show strong sequence similarity with olfactomedin. This similarity is suggestive of a matrix-related function of these glycoproteins in neurons and neurosecretory cells (Danielson, P.E. et al. (1994) J. Neurosci. Res. 38:468-478).

Mac-2 binding protein is a 90-kD serum protein (90K), a secreted glycoprotein isolated from both the human breast carcinoma cell line SK-BR-3, and human breast milk. It specifically binds to a human macrophage-associated lectin, Mac-2. Structurally, the mature protein is 567 amino acids in length and is proceeded by an 18-amino acid leader. There are 16 cysteines and seven potential N-linked glycosylation sites. The first 106 amino acids represent a domain very similar to an ancient protein superfamily defined by a macrophage scavenger receptor cysteine-rich domain (Koths, K. et al. (1993) J. Biol. Chem. 268:14245-14249). 90K is elevated in the serum of subpopulations of AIDS patients and is expressed at varying levels in primary tumor samples and tumor cell lines. Ullrich et al. (1994) have demonstrated that 90K stimulates host defense systems and can induce interleukin-2 secretion. This immune stimulation is proposed to be a result of oncogenic transformation, viral infection or pathogenic invasion (Ullrich, A. et al. (1994) J. Biol. Chem. 269:18401-18407).

Semaphorins are a large group of axonal guidance molecules consisting of at least 30 different

members and are found in vertebrates, invertebrates, and even certain viruses. All semaphorins contain the sema domain which is approximately 500 amino acids in length. Neuropilin, a semaphorin receptor, has been shown to promote neurite outgrowth in vitro. The extracellular region of neuropilins consists of three different domains: CUB, discoidin, and MAM domains. The CUB and the MAM motifs of neuropilin have been suggested to have roles in protein-protein interactions and are thought to be involved in the binding of semaphorins through the sema and the C-terminal domains (reviewed in Raper, J.A. (2000) Curr. Opin. Neurobiol. 10:88-94). Plexins are neuronal cell surface molecules that mediate cell adhesion via a homophilic binding mechanism in the presence of calcium ions. Plexins have been shown to be expressed in the receptors and neurons of particular sensory systems (Ohta, K. et al. (1995) Cell 14:1189-1199). There is evidence that suggests that some plexins function to control motor and CNS axon guidance in the developing nervous system. Plexins, which themselves contain complete semaphorin domains, may be both the ancestors of classical semaphorins and binding partners for semaphorins (Winberg, M.L. et al (1998) Cell 95:903-916).

Human pregnancy-specific beta 1-glycoprotein (PSG) is a family of closely related glycoproteins of molecular weights of 72 KDa, 64KDa, 62KDa, and 54KDa. Together with the carcinoembryonic antigen, they comprise a subfamily within the immunoglobulin superfamily (Plouzek, C.A. and Chou, J.Y. (1991) Endocrinology 129:950-958) Different subpopulations of PSG have been found to be produced by the trophoblasts of the human placenta, and the amnionic and chorionic membranes (Plouzek, C.A. et al. (1993) Placenta 14:277-285).

Autocrine motility factor (AMF) is one of the motility cytokines regulating tumor cell migration; therefore identification of the signaling pathway coupled with it has critical importance. Autocrine motility factor receptor (AMFR) expression has been found to be associated with tumor progression in thymoma (Ohta Y. et al. (2000) Int. J. Oncol. 17:259-264). AMFR is a cell surface glycoprotein of molecular weight 78KDa.

Hormones are secreted molecules that travel through the circulation and bind to specific receptors on the surface of, or within, target cells. Although they have diverse biochemical compositions and mechanisms of action, hormones can be grouped into two categories. One category includes small lipophilic hormones that diffuse through the plasma membrane of target cells, bind to cytosolic or nuclear receptors, and form a complex that alters gene expression. Examples of these molecules include retinoic acid, thyroxine, and the cholesterol-derived steroid hormones such as progesterone, estrogen, testosterone, cortisol, and aldosterone. The second category includes hydrophilic hormones that function by binding to cell surface receptors that transduce signals across the plasma membrane. Examples of such hormones include amino acid derivatives such as

catecholamines (epinephrine, norepinephrine) and histamine, and peptide hormones such as glucagon, insulin, gastrin, secretin, cholecystokinin, adrenocorticotropic hormone, follicle stimulating hormone, luteinizing hormone, thyroid stimulating hormone, and vasopressin. (See, for example, Lodish et al. (1995) Molecular Cell Biology, Scientific American Books Inc., New York, NY, pp. 856-864.)

Pro-opiomelanocortin (POMC) is the precursor polypeptide of corticotropin (ACTH), a hormone synthesized by the anterior pituitary gland, which functions in the stimulation of the adrenal cortex. POMC is also the precursor polypeptide of the hormone beta-lipotropin (beta-LPH). Each hormone includes smaller peptides with distinct biological activities: alpha-melanotropin (alpha-MSH) and corticotropin-like intermediate lobe peptide (CLIP) are formed from ACTH; gamma-lipotropin (gamma-LPH) and beta-endorphin are peptide components of beta-LPH; while beta-MSH is contained within gamma-LPH. Adrenal insufficiency due to ACTH deficiency, resulting from a genetic mutation in exons 2 and 3 of POMC results in an endocrine disorder characterized by early-onset obesity, adrenal insufficiency, and red hair pigmentation (Chretien, M. et al. (1979) Can. J. Biochem. 57:1111-1121; Krude, H. et al. (1998) Nat. Genet. 19:155-157; Online Mendelian Inheritance in Man (OMIM) 176830).

Growth and differentiation factors are secreted proteins which function in intercellular communication. Some factors require oligomerization or association with membrane proteins for activity. Complex interactions among these factors and their receptors trigger intracellular signal transduction pathways that stimulate or inhibit cell division, cell differentiation, cell signaling, and cell motility. Most growth and differentiation factors act on cells in their local environment (paracrine signaling). There are three broad classes of growth and differentiation factors. The first class includes the large polypeptide growth factors such as epidermal growth factor, fibroblast growth factor, transforming growth factor, insulin-like growth factor, and platelet-derived growth factor. The second class includes the hematopoietic growth factors such as the colony stimulating factors (CSFs). Hematopoietic growth factors stimulate the proliferation and differentiation of blood cells such as B-lymphocytes, T-lymphocytes, erythrocytes, platelets, eosinophils, basophils, neutrophils, macrophages, and their stem cell precursors. The third class includes small peptide factors such as bombesin, vasopressin, oxytocin, endothelin, transferrin, angiotensin II, vasoactive intestinal peptide, and bradykinin, which function as hormones to regulate cellular functions other than proliferation.

Growth and differentiation factors play critical roles in neoplastic transformation of cells in vitro and in tumor progression in vivo. Inappropriate expression of growth factors by tumor cells may contribute to vascularization and metastasis of tumors. During hematopoiesis, growth factor misregulation can result in anemias, leukemias, and lymphomas. Certain growth factors such as

interferon are cytotoxic to tumor cells both in vivo and in vitro. Moreover, some growth factors and growth factor receptors are related both structurally and functionally to oncoproteins. In addition, growth factors affect transcriptional regulation of both proto-oncogenes and oncosuppressor genes. (Reviewed in Pimentel, E. (1994) Handbook of Growth Factors, CRC Press, Ann Arbor, MI, pp. 1-9.)

The Slit protein, first identified in Drosophila, is critical in central nervous system midline formation and potentially in nervous tissue histogenesis and axonal pathfinding. Itoh et al. ((1998) Brain Res. Mol. Brain Res. 62:175-186) have identified mammalian homologues of the slit gene (human Slit-1, Slit-2, Slit-3 and rat Slit-1). The encoded proteins are putative secreted proteins containing EGF-like motifs and leucine-rich repeats, both of which are conserved protein-protein interaction domains. Slit-1, -2, and -3 mRNAs are expressed in the brain, spinal cord, and thyroid, respectively (Itoh, A. et al., supra). The Slit family of proteins are indicated to be functional ligands of glypican-1 in nervous tissue and it is suggested that their interactions may be critical in certain stages during central nervous system histogenesis (Liang, Y. et al. (1999) J. Biol. Chem. 274:17885-17892).

Neuropeptides and vasomediators (NP/VM) comprise a large family of endogenous signaling molecules. Included in this family are neuropeptides and neuropeptide hormones such as bombesin, neuropeptide Y, neurotensin, neuromedin N, melanocortins, opioids, galanin, somatostatin, tachykinins, urotensin II and related peptides involved in smooth muscle stimulation, vasopressin, vasoactive intestinal peptide, and circulatory system-borne signaling molecules such as angiotensin, complement, calcitonin, endothelins, formyl-methionyl peptides, glucagon, cholecystokinin and gastrin. NP/VMs can transduce signals directly, modulate the activity or release of other neurotransmitters and hormones, and act as catalytic enzymes in cascades. The effects of NP/VMs range from extremely brief to long-lasting. (Reviewed in Martin, C.R. et al. (1985) Endocrine Physiology, Oxford University Press, New York, NY, pp. 57-62.)

NP/VMs are involved in numerous neurological and cardiovascular disorders. For example, neuropeptide Y is involved in hypertension, congestive heart failure, affective disorders, and appetite regulation. Somatostatin inhibits secretion of growth hormone and prolactin in the anterior pituitary, as well as inhibiting secretion in intestine, pancreatic acinar cells, and pancreatic beta-cells. A reduction in somatostatin levels has been reported in Alzheimer's disease and Parkinson's disease. Vasopressin acts in the kidney to increase water and sodium absorption, and in higher concentrations stimulates contraction of vascular smooth muscle, platelet activation, and glycogen breakdown in the liver. Vasopressin and its analogues are used clinically to treat diabetes insipidus. Endothelin and angiotensin are involved in hypertension, and drugs, such as captopril, which reduce plasma levels of angiotensin, are used to reduce blood pressure (Watson, S. and S. Arkinstall (1994) The G-protein

Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 194; 252; 284; 55; 111).

Neuropeptides have also been shown to have roles in nociception (pain). Vasoactive intestinal peptide appears to play an important role in chronic neuropathic pain. Nociceptin, an endogenous ligand for for the opioid receptor-like 1 receptor, is thought to have a predominantly anti-nociceptive effect, and has been shown to have analgesic properties in different animal models of tonic or chronic pain (Dickinson, T. and Fleetwood-Walker, S.M. (1998) Trends Pharmacol. Sci. 19:346-348).

Other proteins that contain signal peptides include secreted proteins with enzymatic activity. Such activity includes, for example, oxidoreductase/dehydrogenase activity, transferase activity, hydrolase activity, lyase activity, isomerase activity, or ligase activity. For example, matrix metalloproteinases are secreted hydrolytic enzymes that degrade the extracellular matrix and thus play an important role in tumor metastasis, tissue morphogenesis, and arthritis (Reponen, P. et al. (1995) Dev. Dyn. 202:388-396; Firestein, G.S. (1992) Curr. Opin. Rheumatol. 4:348-354; Ray, J.M. and Stetler-Stevenson, W.G. (1994) Eur. Respir. J. 7:2062-2072; and Mignatti, P. and Rifkin, D.B. (1993) Physiol. Rev. 73:161-195). Additional examples are the acetyl-CoA synthetases which activate acetate for use in lipid synthesis or energy generation (Luong, A. et al. (2000) J. Biol. Chem. 275:26458-26466). The result of acetyl-CoA synthetase activity is the formation of acetyl-CoA from acetate and CoA. Acetyl-CoA sythetases share a region of sequence similarity identified as the AMP-binding domain signature. Acetyl-CoA synthetase has been shown to be associated with hypertension (Toh, H. (1991) Protein Seq. Data Anal. 4:111-117; and Iwai, N. et al. (1994) Hypertension 23:375-380).

A number of isomerases catalyze steps in protein folding, phototransduction, and various anabolic and catabolic pathways. One class of isomerases is known as peptidyl-prolyl *cis-trans* isomerases (PPIases). PPIases catalyze the *cis* to *trans* isomerization of certain proline imidic bonds in proteins. Two families of PPIases are the FK506 binding proteins (FKBPs), and cyclophilins (CyPs). FKBPs bind the potent immunosuppressants FK506 and rapamycin, thereby inhibiting signaling pathways in T-cells. Specifically, the PPIase activity of FKBPs is inhibited by binding of FK506 or rapamycin. There are five members of the FKBP family which are named according to their calculated molecular masses (FKBP12, FKBP13, FKBP25, FKBP52, and FKBP65), and localized to different regions of the cell where they associate with different protein complexes (Coss, M. et al. (1995) J. Biol. Chem. 270:29336-29341; Schreiber, S.L. (1991) Science 251:283-287).

The peptidyl-prolyl isomerase activity of CyP may be part of the signaling pathway that leads to T-cell activation. CyP isomerase activity is associated with protein folding and protein trafficking, and may also be involved in assembly/disassembly of protein complexes and regulation of protein

activity. For example, in *Drosophila*, the CyP NinaA is required for correct localization of rhodopsins, while a mammalian CyP (Cyp40) is part of the Hsp90/Hsc70 complex that binds steroid receptors. The mammalian CypA has been shown to bind the *gag* protein from human immunodeficiency virus 1 (HIV-1), an interaction that can be inhibited by cyclosporin. Since cyclosporin has potent anti-HIV-1 activity, CypA may play an essential function in HIV-1 replication. Finally, Cyp40 has been shown to bind and inactivate the transcription factor c-Myb, an effect that is reversed by cyclosporin. This effect implicates CyPs in the regulation of transcription, transformation, and differentiation (Bergsma, D.J. et al (1991) J. Biol. Chem. 266:23204-23214; Hunter, T. (1998) Cell 92:141-143; and Leverson, J.D. and Ness, S.A. (1998) Mol. Cell. 1:203-211).

Gamma-carboxyglutamic acid (Gla) proteins rich in proline (PRGPs) are members of a family of vitamin K-dependent single-pass integral membrane proteins. These proteins are characterized by an extracellular amino terminal domain of approximately 45 amino acids rich in Gla. The intracellular carboxyl terminal region contains one or two copies of the sequence PPXY, a motif present in a variety of proteins involved in such diverse cellular functions as signal transduction, cell cycle progression, and protein turnover (Kulman, J.D. et al. (2001) Proc. Natl. Acad. Sci. USA 98:1370-1375). The process of post-translational modification of glutamic residues to form Gla is Vitamin K-dependent carboxylation. Proteins which contain Gla include plasma proteins involved in blood coagulation. These proteins are prothrombin, proteins C, S, and Z, and coagulation factors VII, IX, and X. Osteocalcin (bone-Gla protein, BGP) and matrix Gla-protein (MGP) also contain Gla (Friedman, P.A. and C.T. Przysiecki (1987) Int. J. Biochem. 19:1-7; C. Vermeer (1990) Biochem. J. 266:625-636).

The discovery of new full-length human molecules for disease detection and treatment, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, developmental, neurological, and cardiovascular disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of full-length human molecules for disease detection and treatment.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, full-length human molecules for disease detection and treatment, referred to collectively as "MDDT" and individually as "MDDT-1," "MDDT-2," "MDDT-3," "MDDT-4," "MDDT-5," "MDDT-6," "MDDT-7," "MDDT-8," "MDDT-9," "MDDT-10," "MDDT-11," "MDDT-12," "MDDT-13," "MDDT-14," "MDDT-15," "MDDT-16," "MDDT-

17," "MDDT-18," "MDDT-19," and "MDDT-20." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-20.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-20. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:21-40.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional MDDT, comprising

administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional MDDT, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity

of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be

used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"MDDT" refers to the amino acid sequences of substantially purified MDDT obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of MDDT. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of MDDT either by directly interacting with MDDT or by acting on components of the biological pathway in which MDDT participates.

An "allelic variant" is an alternative form of the gene encoding MDDT. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding MDDT include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as MDDT or a polypeptide with at least one functional characteristic of MDDT. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding MDDT, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding MDDT. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent MDDT. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of MDDT is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may

include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of MDDT. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of MDDT either by directly interacting with MDDT or by acting on components of the biological pathway in which MDDT participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind MDDT polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No.

5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic MDDT, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid

sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding MDDT or fragments of MDDT may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr

Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of MDDT or the polynucleotide encoding MDDT which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid

residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:21-40 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:21-40, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:21-40 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:21-40 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:21-40 and the region of SEQ ID NO:21-40 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-20 is encoded by a fragment of SEQ ID NO:21-40. A fragment of SEQ ID NO:1-20 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-20. For example, a fragment of SEQ ID NO:1-20 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-20. The precise length of a fragment of SEQ ID NO:1-20 and the region of SEQ ID NO:1-20 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of

molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a

length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment

length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance,

sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of MDDT which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of MDDT which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of MDDT. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of MDDT.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the

antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an MDDT may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of MDDT.

"Probe" refers to nucleic acid sequences encoding MDDT, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR

Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence.

Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing MDDT, nucleic acids encoding MDDT, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater

sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human full-length human molecules for disease detection and treatment (MDDT), the polynucleotides encoding MDDT, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, developmental, neurological, and cardiovascular disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by

BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are full-length human molecules for disease detection and treatment. For example, SEQ ID NO:3 is 96% identical, from residue M1 to residue V725, to rat corneal wound healing related protein (GenBank ID g8926320) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from BLAST analyses provide further corroborative evidence that SEQ ID NO:3 is a human full-length molecule for disease detection and treatment. In an alternative example, SEQ ID NO:7 is 24% identical, from residue E214 to residue T735, to corn calmodulin-binding protein MPCBP (GenBank ID g10086260) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.2e-21, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 also contains TPR domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis provide further corroborative evidence that SEQ ID NO:7 is a full-length human molecule for disease

detection and treatment. In an alternative example, SEQ ID NO:10 is 63% identical, from residue P239 to residue V1461, to rat periaxin (GenBank ID g505297) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 also contains a PDZ domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST analyses provide further corroborative evidence that SEQ ID NO:10 is a periaxin. In an alternative example, SEQ ID NO:14 is 36% identical, from residue Y20 to residue V203, to a putative phosphatidylinositol-4-phosphate 5-kinase from thale cress (GenBank ID g2739367) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.0e-25, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:14 also contains a MORN motif as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST analyses provide further corroborative evidence that SEQ ID NO:14 is a kinase. SEQ ID NO:1-2, SEQ ID NO:4-6, SEQ ID NO:8-9, SEQ ID NO:11-13, and SEQ ID NO:15-20 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-20 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:21-40 or that distinguish between SEQ ID NO:21-40 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the

ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).

INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses MDDT variants. A preferred MDDT variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the MDDT amino acid sequence, and which contains at least one functional or structural characteristic of MDDT.

The invention also encompasses polynucleotides which encode MDDT. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40, which encodes MDDT. The polynucleotide sequences of SEQ ID NO:21-40, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding MDDT. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding MDDT. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:21-40. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of MDDT.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding MDDT. A splice variant may have portions which have significant

sequence identity to the polynucleotide sequence encoding MDDT, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding MDDT over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding MDDT. For example, a polynucleotide comprising a sequence of SEQ ID NO:21 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:39 and a polynucleotide comprising a sequence of SEQ ID NO:34 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:40. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of MDDT.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding MDDT, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring MDDT, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode MDDT and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring MDDT under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding MDDT or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding MDDT and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode MDDT and MDDT derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the

synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding MDDT or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:21-40 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding MDDT may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and

ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode MDDT may be cloned in recombinant DNA molecules that direct expression of MDDT, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express MDDT.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter MDDT-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic

oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of MDDT, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding MDDT may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, MDDT itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of MDDT, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active MDDT, the nucleotide sequences encoding MDDT or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding MDDT. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding MDDT. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding MDDT and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding MDDT and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding MDDT. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and

Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding MDDT. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding MDDT can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding MDDT into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of MDDT are needed, e.g. for the production of antibodies, vectors which direct high level expression of MDDT may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of MDDT. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of MDDT. Transcription of sequences encoding MDDT may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated

transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding MDDT may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses MDDT in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of MDDT in cell lines is preferred. For example, sequences encoding MDDT can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which

alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding MDDT is inserted within a marker gene sequence, transformed cells containing sequences encoding MDDT can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding MDDT under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding MDDT and that express MDDT may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of MDDT using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on MDDT is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding MDDT include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding MDDT, or any fragments thereof, may be cloned into a vector

for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding MDDT may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode MDDT may be designed to contain signal sequences which direct secretion of MDDT through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding MDDT may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric MDDT protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of MDDT activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion

proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the MDDT encoding sequence and the heterologous protein sequence, so that MDDT may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled MDDT may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

MDDT of the present invention or fragments thereof may be used to screen for compounds that specifically bind to MDDT. At least one and up to a plurality of test compounds may be screened for specific binding to MDDT. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of MDDT, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which MDDT binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express MDDT, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing MDDT or cell membrane fractions which contain MDDT are then contacted with a test compound and binding, stimulation, or inhibition of activity of either MDDT or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with MDDT, either in solution or affixed to a solid support, and detecting the binding of MDDT to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

MDDT of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of MDDT. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for MDDT activity, wherein MDDT is combined with at least one test compound, and the activity of MDDT in the presence of a test compound is compared with the activity of MDDT in the absence of the test compound. A change in the activity of MDDT in the presence of the test compound is indicative of a compound that modulates the activity of MDDT. Alternatively, a test compound is combined with an in vitro or cell-free system comprising MDDT under conditions suitable for MDDT activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of MDDT may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding MDDT or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding MDDT may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding MDDT can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region

of a polynucleotide encoding MDDT is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress MDDT, e.g., by secreting MDDT in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of MDDT and full-length human molecules for disease detection and treatment. In addition, examples of tissues expressing MDDT can be found in Table 6. Therefore, MDDT appears to play a role in cell proliferative, autoimmune/inflammatory, developmental, neurological, and cardiovascular disorders. In the treatment of disorders associated with increased MDDT expression or activity, it is desirable to decrease the expression or activity of MDDT. In the treatment of disorders associated with decreased MDDT expression or activity, it is desirable to increase the expression or activity of MDDT.

Therefore, in one embodiment, MDDT or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic

lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, hematopoietic cancer including lymphoma, leukemia, and myeloma, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease,

degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery.

In another embodiment, a vector capable of expressing MDDT or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified MDDT in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of MDDT may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT including, but not limited to, those listed above.

In a further embodiment, an antagonist of MDDT may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of MDDT. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, developmental, neurological, and cardiovascular disorders described above. In one aspect, an antibody which specifically binds MDDT may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express MDDT.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding MDDT may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of MDDT including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic

efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of MDDT may be produced using methods which are generally known in the art. In particular, purified MDDT may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind MDDT. Antibodies to MDDT may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with MDDT or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to MDDT have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of MDDT amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to MDDT may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate

antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce MDDT-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for MDDT may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between MDDT and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering MDDT epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for MDDT. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of MDDT-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple MDDT epitopes, represents the average affinity, or avidity, of the antibodies for MDDT. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular MDDT epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the MDDT-

antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of MDDT, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of MDDT-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding MDDT, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding MDDT. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding MDDT. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding MDDT may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in MDDT expression or regulation causes disease, the expression of MDDT from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in MDDT are treated by constructing mammalian expression vectors encoding MDDT and introducing these vectors by mechanical means into MDDT-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of MDDT include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). MDDT may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter

(e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding MDDT from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to MDDT expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding MDDT under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998)

Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding MDDT to cells which have one or more genetic abnormalities with respect to the expression of MDDT. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding MDDT to target cells which have one or more genetic abnormalities with respect to the expression of MDDT. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing MDDT to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding MDDT to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based

on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for MDDT into the alphavirus genome in place of the capsid-coding region results in the production of a large number of MDDT-coding RNAs and the synthesis of high levels of MDDT in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of MDDT into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding MDDT.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides,

corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding MDDT. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding MDDT. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased MDDT expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding MDDT may be therapeutically useful, and in the treatment of disorders associated with decreased MDDT expression or activity, a compound which specifically promotes expression of the polynucleotide encoding MDDT may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method

commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding MDDT is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding MDDT are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding MDDT. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient.

Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of MDDT, antibodies to MDDT, and mimetics, agonists, antagonists, or inhibitors of MDDT.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising MDDT or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, MDDT or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example MDDT or fragments thereof, antibodies of MDDT, and agonists, antagonists or inhibitors of MDDT,

which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind MDDT may be used for the diagnosis of disorders characterized by expression of MDDT, or in assays to monitor patients being treated with MDDT or agonists, antagonists, or inhibitors of MDDT. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for MDDT include methods which utilize the antibody and a label to detect MDDT in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring MDDT, including ELISAs, RIAs, and FACS, are known

in the art and provide a basis for diagnosing altered or abnormal levels of MDDT expression. Normal or standard values for MDDT expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to MDDT under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of MDDT expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding MDDT may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of MDDT may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of MDDT, and to monitor regulation of MDDT levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding MDDT or closely related molecules may be used to identify nucleic acid sequences which encode MDDT. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding MDDT, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the MDDT encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:21-40 or from genomic sequences including promoters, enhancers, and introns of the MDDT gene.

Means for producing specific hybridization probes for DNAs encoding MDDT include the cloning of polynucleotide sequences encoding MDDT or MDDT derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding MDDT may be used for the diagnosis of disorders associated with expression of MDDT. Examples of such disorders include, but are not limited to, a

cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, hematopoietic cancer including lymphoma, leukemia, and myeloma, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis

pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery. The polynucleotide sequences encoding MDDT may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered MDDT expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding MDDT may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding MDDT may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control

sample then the presence of altered levels of nucleotide sequences encoding MDDT in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of MDDT, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding MDDT, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding MDDT may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding MDDT, or a fragment of a polynucleotide complementary to the polynucleotide encoding MDDT, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding MDDT may be used to detect single nucleotide polymorphisms (SNPs). SNPs are

substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding MDDT are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of MDDT include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C.

et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, MDDT, fragments of MDDT, or antibodies specific for MDDT may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating

and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for MDDT to quantify the levels of MDDT expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of

each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding MDDT may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic

map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding MDDT on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, MDDT, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between MDDT and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with MDDT, or fragments thereof, and washed. Bound MDDT is then detected by methods well known in the art. Purified MDDT can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding MDDT specifically compete with a test compound for binding MDDT. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with MDDT.

In additional embodiments, the nucleotide sequences which encode MDDT may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/268,117, U.S. Ser. No. 60/269,618, U.S. Ser. No. 60/271,118, U.S. Ser. No. 60/274,436, U.S. Ser. No. 60/274,486, U.S. Ser. No. 60/344,229, and Attorney Docket No. PF-1352 P filed February 1, 2002, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic

oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM; and HMM-based protein domain databases such as SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software

Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:21-40. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative full-length human molecules for disease detection and treatment were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode full-length human molecules for disease detection and treatment, the encoded polypeptides were analyzed by querying against PFAM models for full-length human molecules for disease detection and treatment. Potential full-length human molecules for disease detection and treatment were also identified by homology to Incyte cDNA sequences that had been annotated as full-length human molecules for disease detection and treatment. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the

sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases

using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of MDDT Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:21-40 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:21-40 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel

(1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum } \{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding MDDT are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following

disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding MDDT. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of MDDT Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the

sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in MDDT Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:21-40 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files

in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:21-40 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and

compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one

with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly

larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission

spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XII. Complementary Polynucleotides

Sequences complementary to the MDDT-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring MDDT. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of MDDT. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the MDDT-encoding transcript.

XIII. Expression of MDDT

Expression and purification of MDDT is achieved using bacterial or virus-based expression systems. For expression of MDDT in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express MDDT upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of MDDT in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding MDDT by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, MDDT is synthesized as a fusion protein with, e.g., glutathione S-

transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from MDDT at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified MDDT obtained by these methods can be used directly in the assays shown in Examples XVII and XVIII, where applicable.

XIV. Functional Assays

MDDT function is assessed by expressing the sequences encoding MDDT at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of MDDT on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding MDDT and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding MDDT and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of MDDT Specific Antibodies

MDDT substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the MDDT amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-MDDT activity by, for example, binding the peptide or MDDT to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring MDDT Using Specific Antibodies

Naturally occurring or recombinant MDDT is substantially purified by immunoaffinity chromatography using antibodies specific for MDDT. An immunoaffinity column is constructed by covalently coupling anti-MDDT antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing MDDT are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of MDDT (e.g., high ionic strength

buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/MDDT binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and MDDT is collected.

XVII. Identification of Molecules Which Interact with MDDT

MDDT, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled MDDT, washed, and any wells with labeled MDDT complex are assayed. Data obtained using different concentrations of MDDT are used to calculate values for the number, affinity, and association of MDDT with the candidate molecules.

Alternatively, molecules interacting with MDDT are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

MDDT may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of MDDT Activity

An assay for growth stimulating or inhibiting activity of MDDT measures the amount of DNA synthesis in Swiss mouse 3T3 cells (McKay, I. and Leigh, I., eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York, NY). In this assay, varying amounts of MDDT are added to quiescent 3T3 cultured cells in the presence of [^{3}H]thymidine, a radioactive DNA precursor. MDDT for this assay can be obtained by recombinant means or from biochemical preparations. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold MDDT concentration range is indicative of growth modulating activity. One unit of activity per milliliter is defined as the concentration of MDDT producing a 50% response level, where 100% represents maximal incorporation of [^{3}H]thymidine into acid-precipitable DNA .

Alternatively, an assay for MDDT activity measures the stimulation or inhibition of neurotransmission in cultured cells. Cultured CHO fibroblasts are exposed to MDDT. Following endocytic uptake of MDDT, the cells are washed with fresh culture medium, and a whole cell voltage-clamped Xenopus myocyte is manipulated into contact with one of the fibroblasts in MDDT-free

medium. Membrane currents are recorded from the myocyte. Increased or decreased current relative to control values are indicative of neuromodulatory effects of MDDT (Morimoto, T. et al. (1995) Neuron 15:689-696).

Alternatively, an assay for MDDT activity measures the amount of MDDT in secretory, membrane-bound organelles. Transfected cells as described above are harvested and lysed. The lysate is fractionated using methods known to those of skill in the art, for example, sucrose gradient ultracentrifugation. Such methods allow the isolation of subcellular components such as the Golgi apparatus, ER, small membrane-bound vesicles, and other secretory organelles. Immunoprecipitations from fractionated and total cell lysates are performed using MDDT-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The concentration of MDDT in secretory organelles relative to MDDT in total cell lysate is proportional to the amount of MDDT in transit through the secretory pathway.

Alternatively, AMP binding activity is measured by combining MDDT with ^{32}P-labeled AMP. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to remove unbound label. The radioactivity retained in the gel is proportional to MDDT activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1419725	1	1419725CD1	21	1419725CB1
628613	2	628613CD1	22	628613CB1
7111920	3	7111920CD1	23	7111920CB1
3072268	4	3072268CD1	24	3072268CB1
5519523	5	5519523CD1	25	5519523CB1
1760208	6	1760208CD1	26	1760208CB1
1900132	7	1900132CD1	27	1900132CB1
7487551	8	7487551CD1	28	7487551CB1
1871014	9	1871014CD1	29	1871014CB1
2903166	10	2903166CD1	30	2903166CB1
1723804	11	1723804CD1	31	1723804CB1
7736769	12	7736769CD1	32	7736769CB1
7492451	13	7492451CD1	33	7492451CB1
4650669	14	4650669CD1	34	4650669CB1
7485268	15	7485268CD1	35	7485268CB1
2112995	16	2112995CD1	36	2112995CB1
1613452	17	1613452CD1	37	1613452CB1
55061615	18	55061615CD1	38	55061615CB1
7503435	19	7503435CD1	39	7503435CB1
7504149	20	7504149CD1	40	7504149CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability score	Annotation
3	7111920CD1	g8926320	0.0	[Rattus norvegicus] corneal wound healing related protein (Yi, X.J. et al. (2000) Curr. Eye Res. 20:430-440)
4	3072268CD1	g12002207	0.0	chymotrypsin-like protein [Homo sapiens]
		g6581056	8.6E-183	[Homo sapiens] CHORD containing protein-1 (Shirasu, K. et al. (1999) Cell 99:355-366)
5	5519523CD1	g15487240	0.0	putative autophagy-related cysteine endopeptidase 2 [Homo sapiens]
6	1760208CD1	g17907795	0.0	TGF-beta induced apotosis protein 3 [Homo sapiens]
7	1900132CD1	g10086260	1.2E-21	[Zea mays] calmodulin-binding protein MPCBP (Safadi,F. et al. (2000) J. Biol. Chem. 275:35457-35470)
8	7487551CD1	g520740	8.2E-84	[Homo sapiens] olfactory marker protein (Buiakova, O.I. et al. (1994) Genomics 20:452-462)
10	2903166CD1	g505297	0.0	[Rattus norvegicus] periaxin (Gillespie, C.S. et al. (1994) Neuron 12:497-508)
12	7736769CD1	g10636484	6.3E-113	[Homo sapiens] polyglutamine-containing protein (Rampazzo, A. et al. (2000) Biochem. Biophys. Res. Commun. 278:766-774)
13	7492451CD1	g2879800	2.2E-21	[Schizosaccharomyces pombe] phenylalanyl-trna synthetase, alpha chain, cytoplasmic
14	4650669CD1	g2739367	2.0E-25	[Arabidopsis thaliana] putative phosphatidylinositol-4-phosphate 5-kinase
15	7485268CD1	g13274531	1.0E-64	complement-c1q tumor necrosis factor-related protein [Homo sapiens]
16	2112995CD1	g3126975	2.5E-263	[Mus musculus] retinoic acid-responsive protein; STRA6 (Bouillet, P. et al. (1995) Dev. Biol. 170:420-433)
18	55061615CD1	g10432393	2.5E-206	dJ947L8.1.8 (novel Sushi (SCR repeat) domain protein) [Homo sapiens]

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability score	Annotation
20	7504149CD1	g13925629	1.4E-18	[Arabidopsis thaliana] phosphatidylinositol-4-phosphate 5-kinase
		692644|Tsga2	6.0E-104	[Mus musculus] Testis-specific protein, expressed during spermatogenesis. Taketo, M. M. et al. (1997) Genomics 46:138-142.

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1419725CD1	198	S26 S30 T18 T108 T185		Signal peptide: M48-A72	HMMER
2	628613CD1	385	S33 S75 S188 S264 S305 S342 S354 T127 T346 T361	N167 N170 N254 N257 N333 N377	Hypothetical protein KIAA0009: PD128946: L24-G175, L118-Q359	BLAST-PRODOM
3	7111920CD1	725	S6 S12 S17 S61 S187 S210 S406 S506 S514 S538 S559 S560 S705 T94 T101 T118 T251 T255 T289 T290 T338 T459 T563 Y328 Y646	N287 N344	Transmembrane domain: A232-A249 N-terminus is non-cytosolic	TMAP
					T23B12.4 protein PD148039: G292-R682, M18-L247, E671-A698	BLAST-PRODOM
					Glucose repressible protein MAK10 PD147352: V30-F182, T490-A581, K566-E639	BLAST-PRODOM
4	3072268CD1	332	S66 S110 S125 S137 S156 S171 S200 S250 S255 T18 T47 T48 T80 T116 T199 T219 T237 T298 T303	N260	Signal peptide: M1-G62	SPScan
5	5519523CD1	402	S10 S54 S66 S145 T44 T60 T199 T289 T298 T377	N212 N296	Protein F6E13.27, ZK792.1, URE2SSU72 intergenic region PD152705: Q213-W337, P27-L162	BLAST-PRODOM
6	1760208CD1	589	S45 S124 S179 S233 S308 S322 S396 S398 S493 S500 S522 S573 S582	N16 N419	Signal peptide: M1-S30	SPScan
					Similarity to rat mitochondrial capsule selenoprotein PD144344: R205-E305	BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	1900132CD1	741	S100 S187 S251 S311 S386 S409 S523 S528 S549 S564 S643 S679 T72 T205 T355 T503 T533 T544 T694 T721 Y176 Y435	N132 N516 N692	TPR Domain: H628-H661, H696-S729, A447-N480, V662-A695, F295-D328, A594-S627, H413-D446	HMMER-PFAM
					Transmembrane domain: K289-L305 N-terminus is non-cytosolic	TMAP
					Kinesin light chain repeat proteins BL01160: E646-S693, D445-A473	BLIMPS-BLOCKS
8	7487551CD1	227	S95 T208	N120	Signal peptide: M1-A66	SPScan
					Olfactory marker protein, neuronal specific, PD022055: P70-F224	BLAST-PRODOM
9	1871014CD1	261	S105 T25 T231 T257 Y140	N193	Leucine zipper pattern: L200-L221	MOTIFS
10	2903166CD1	1461	S7 S58 S67 S113 S399 S430 S828 S928 S1004 Y77 S1082 S1275 S1328 S1339 S1351 S1368 S1407 S1418 T419 T787 T1130		PDZ domain (also known as DHR or GLGF): E18-T99	HMMER-PFAM
					Periaxin repeat: PD041976: R1070-E1342 PD018116: K136-R404 PD021686: M1-V135 PD155663: V577-P668	BLAST-PRODOM
					Neurofilament, triplet: DM04498|P12036|434-1019: G341-D842	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	1723804CD1	657	S18 S30 S55 S79 S84 S203 S332 S468 S473 S570 S576 S579 S580 S621 T6 T52 T62 T150 T173 T234 T274 T275 T323 T484 T534 T593 Y490	N77 N97 N106 N283 N574	Poly(ADP-ribose) polymerase zinc finger domain proteins BL00347: S473-I524, N546-S600	BLIMPS-BLOCKS
12	7736769CD1	587	S11 S15 S87 S135 S163 S360 S409 S455 S492 S526 S537 T55 T304 T353 T382 T413 T450 T451	N193 N345 N410	Zinc finger, C3HC4 type (RING finger): C506-S551	HMMER-PFAM
					M04G12.1 protein PD138197: R452-S587	BLAST-PRODOM
					Cytochrome c family heme-binding site signature: C506-E511	MOTIFS
13	7492451CD1	583	S9 S62 S203 S253 S275 S280 S431 S467 S518 S520 S561 S568 T322 T465 T475 T476 T510 T522 Y509	N135 N159	Leucine rich repeat: S203-P225, A100-G121, Q130-P153, R154-A177, Q76-P99, L51-P75, K226-Q250, L180-A202	HMMER-PFAM
					Signal peptide: M1-A37	SPScan
					Phenylalanyl-tRNA synthetase, ligase subunit PD025378: V325-E505	BLAST-PRODOM
					Leucine zipper pattern: L134-L155	MOTIFS
14	4650669CD1	309	S88 S243 S297 T189 T229	N110	MORN motif: Y67-R89, Y90-T112, Y44-R66, Y113-K136, Y159-E181, Y20-T43	HMMER-PFAM
					Phosphatidyl inositol-4-phosphate 5-kinase PD149995: E8-H183, Y20-M191	BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	7485268CD1	252	S43 S84 S197 Y231		Signal peptide: M1-A17	SPScan
					Signal peptide: M1-P22, M1-R24, M1-P25, M1-R31	HMMER
					C1q domain: A118-V246	HMMER-PFAM
					Transmembrane domains: A153-K176, V201-M221	TMAP
					C1q domain proteins BL01113: G88-R114, A135-V170, V201-R220, I239-P248	BLIMPS-BLOCKS
					Complement C1Q domain signature PR00007: S129-R155, F156-Y175, V201-F222, L237-K247	BLIMPS-PRINTS
					C1Q domain: DM00777|P02745|65-244: L68-A250 DM00777|Q06576|37-214: G70-V246 DM00777|P98085|222-418: K69-P248 DM00777|Q02105|71-245: K69-P248	BLAST-DOMO
					Cell attachment sequence: R49-D51	MOTIFS
16	2112995CD1	667	S89 S232 S245 S605 T266 T387 T505 T530 T565	N8	Transmembrane domains: P49-Q75, D97-L117, R143-A165, I200-V228, L294-I322, K356-M382, A429-V457, E470-F494, N506-L534 N-terminus is cytosolic	TMAP
					Retinoic acid responsive protein: PD145028: W77-P667 PD051615: M1-C55	BLAST-PRODOM
					ATP/GTP-binding site motif A (P-loop): A132-T139	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	1613452CD1	657	S36 S68 S103 S143 S321 S410 S590 T45 T71 T119 T136 T163 T221 T265 T271 T276 T293 T319 T401 T402 T518 T577 T607	N443	Similarity to myosin light chain PD146444: S28-L656	BLAST-PRODOM
					Hypothetical 97.0 kD protein PD148168: R37-D469	BLAST-PRODOM
					Cell attachment sequence: R76-D78	MOTIFS
18	55061615CD1	1958	T688 T827 S28 S50 T72 S86 S266 S439 S758 S929 T992 S1011 S1060 T1071 S1113 S1225 S1259 T1329 S1562 T1607 S1660 S1672 S1720 S42 T77 T98 S125 T156 T250 S445 S684 S723 T822 T974 S1016 T1052 T1188 T1357 S1426 S1586 T1667 S1773 S1792 S1803 T1891 S1906 S1925 Y1476 Y1629	N40 N60 N76 N275 N520 N662 N807 N820 N897 N1033 N1206 N1211 N1245 N1416 N1452 N1771 N1896	Signal peptide: M1-F25	SPScan
					CUB domain: C327-Y432, C817-Y922, C501-L595, C989-Y1094, C153-F259, T2-F85, C1377-Y1485, C1203-F1308	HMMER-PFAM
					Sushi domain (SCR repeat): C1839-C1892, C1756-C1809, C1673-C1726, C1144-C1199, C930-C985, C93-C149, C1316-C1373, C1608-C1668, C1537-C1594, C440-C497, C267-C323, C756-C813, Y1476-C1532, C1897-C1955	HMMER-PFAM
					Transmembrane domains: K356-Y384, C1015-I1043, D1228-L1246, L1292-L1313 N-terminus is non-cytosolic	TMAP
					EGF-like domain, glycoprotein PD000165: C327-Y432, T1384-Y1485, C989-Y1094, C817-Y922	BLAST-PRODOM
					Protein F36H2.3A F36H2.3B PD004794: G1494-C1942,	BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					SUSHI repeat: DM04887|P16581|1-609: S1496-L1733 DM04887|P33730|1-610: S1486-L1733 DM04887|P27113|1-551: F1499-C1726	BLAST-DOMO
					C1R/C1S repeat: DM00162|P98069|418-529: A325-Y432	BLAST-DOMO
19	7503435CD1	100	S26 S30 T18		signal_cleavage: M1-A62	SPSCAN
20	7504149CD1	271	S50 S205 S259 T151 T191	N72	MORN repeat: Y29-R51, Y52-T74, Y75-S97, Y121-E143	HMMER_PFAM
					PROTEIN PHOSPHATIDYLINOSITOL-4-PHOSPHATE 5-KINASE PUTATIVE T22C1.7 ISOLOG ATPIP5K1 T4C15.16 PD149995: E12-H145, N25-G146, E12-R149	BLAST_PRODOM

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
21/ 1419725CB1/ 1506	1-241, 1-247, 1-349, 1-422, 1-491, 1-634, 1-1501, 7-266, 11-666, 18-523, 140-369, 323-607, 334-754, 341-831, 345-886, 354-1008, 359-944, 365-747, 367-852, 374-874, 374-881, 376-622, 383-747, 385-691, 387-930, 401-678, 410-790, 437-831, 446-930, 487-719, 487-1042, 518-728, 537-1149, 542-1087, 592-1127, 625-1311, 649-955, 682-1277, 696-1011, 721-902, 759-1036, 778-1326, 835-867, 899-1506, 909-1242, 976-1506, 988-1506, 991-1506, 1037-1343, 1040-1471, 1068-1406, 1085-1506, 1087-1471, 1107-1506, 1157-1385, 1180-1471, 1203-1506, 1211-1471, 1216-1471, 1225-1471, 1274-1420, 1274-1504, 1274-1506, 1275-1506, 1360-1506, 1402-1498, 1420-1506
22/ 628613CB1/ 1565	1-271, 1-501, 69-309, 69-733, 70-500, 70-673, 82-338, 82-630, 95-749, 102-325, 102-445, 105-325, 109-764, 110-362, 110-533, 123-377, 203-953, 321-926, 349-942, 377-884, 394-1130, 433-1018, 464-1142, 486-1169, 490-1045, 588-939, 635-1230, 645-879, 645-1209, 646-1113, 672-867, 672-1027, 691-1230, 761-1372, 831-1073, 860-1146, 932-1541, 1097-1565, 1112-1556, 1116-1565, 1128-1560, 1154-1555, 1166-1563, 1169-1554, 1171-1554, 1173-1560, 1179-1556, 1183-1458, 1209-1480, 1209-1550, 1209-1565, 1225-1560, 1243-1561, 1256-1556, 1346-1560, 1463-1555
23/ 7111920CB1/ 2488	1-100, 1-146, 1-572, 3-146, 13-140, 97-725, 241-760, 431-593, 599-1306, 959-1855, 965-1273, 1642-2322, 1657-1917, 1657-2214, 1657-2220, 1662-2347, 1672-1972, 1674-2046, 1729-2013, 1734-2016, 1760-1985, 1760-2202, 1763-2017, 1763-2051, 1763-2367, 1763-2488, 1767-2039, 1767-2121, 1778-2039, 1783-2024, 1786-2073, 1814-2084, 2070-2124, 2086-2123, 2151-2218
24/ 3072268CB1/ 2647	1-494, 14-276, 47-325, 56-322, 56-476, 56-523, 56-572, 56-600, 59-305, 67-317, 69-361, 69-591, 70-321, 79-341, 86-331, 86-362, 90-327, 90-525, 90-595, 90-645, 93-354, 94-234, 94-265, 94-333, 94-337, 94-355, 94-360, 94-525, 100-391, 100-398, 108-297, 109-420, 112-350, 112-568, 112-759, 115-357, 160-504, 180-811, 225-527, 272-538, 342-588, 355-654, 389-645, 419-1027, 422-654, 457-958, 474-750, 504-1140, 557-825, 591-775, 604-650, 661-819, 744-994, 837-1102, 839-1105, 840-1094, 885-1023, 951-1193, 977-1251, 1012-1320, 1051-1332, 1063-1323, 1084-1511, 1140-1360, 1226-1507, 1247-1525, 1285-1543, 1307-1529, 1358-1616, 1358-1809, 1358-1829, 1370-1547, 1378-1636, 1381-1683, 1432-1670, 1460-1705, 1472-1762, 1499-1719, 1519-1716, 1579-2087, 1621-2088, 1643-2087, 1651-1904, 1651-2074,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
24/ 3072268CB1/ 2647(continued)	1652-2086, 1661-2086, 1694-1887, 1726-1968, 1740-2229, 1743-2226, 1827-2076, 1827-2079, 1827-2125, 1838-2047, 1838-2086, 1881-2169, 1950-2225, 2068-2511, 2080-2318, 2114-2348, 2114-2377, 2116-2575, 2116-2647
25/ 5519523CB1/ 2337	1-241, 170-285, 170-345, 170-368, 170-390, 170-428, 170-435, 170-436, 170-442, 170-453, 170-454, 170-624, 170-638, 170-663, 173-315, 173-405, 173-416, 173-545, 212-285, 245-455, 245-834, 282-480, 353-614, 356-455, 445-750, 501-763, 501-972, 534-778, 802-1118, 842-1073, 842-1075, 842-1077, 842-1080, 842-1082, 842-1087, 842-1090, 842-1445, 844-1091, 844-1095, 844-1417, 845-1071, 845-1075, 845-1092, 845-1100, 845-1116, 917-1181, 945-1184, 968-1123, 968-1433, 968-1503, 1032-1240, 1065-1520, 1106-1373, 1123-1377, 1219-1815, 1235-1410, 1235-1796, 1246-1495, 1264-1542, 1287-1434, 1290-1486, 1316-1580, 1318-1494, 1356-1924, 1414-1826, 1433-1824, 1434-1721, 1450-1745, 1474-2030, 1500-1745, 1513-1801, 1513-1986, 1536-1795, 1536-1805, 1601-1817, 1612-1955, 1637-2072, 1637-2085, 1637-2094, 1644-2295, 1649-1827, 1671-2304, 1674-2294, 1674-2304, 1683-1823, 1686-2300, 1688-2337, 1738-2085, 1797-2069, 1817-2027, 1830-2073, 1878-2307, 1889-2307, 1893-2072, 1896-2151, 1897-2307, 1898-2150, 1916-2307, 1928-2307, 1929-2307, 1931-2307, 1934-2307, 1948-2307, 1961-2307, 1971-2307, 1984-2253, 2024-2307, 2043-2307, 2076-2307
26/ 1760208CB1/ 3141	1-203, 1-365, 1-534, 1-603, 1-708, 1-773, 17-779, 88-542, 115-556, 115-561, 115-571, 115-576, 115-603, 166-692, 244-725, 282-544, 289-536, 289-688, 335-1092, 350-972, 352-1193, 367-659, 377-642, 385-1054, 403-1159, 423-1176, 430-1031, 443-1167, 450-1227, 489-1227, 490-732, 546-985, 578-1114, 701-1299, 727-1251, 765-1367, 792-1229, 811-1324, 885-1008, 885-1375, 886-1524, 904-1108, 910-1063, 943-1393, 985-1199, 993-1591, 1036-1561, 1126-1597, 1168-1674, 1169-1469, 1199-1455, 1217-1503, 1330-1555, 1338-1546, 1348-1606, 1350-1936, 1367-1623, 1368-1626, 1379-1968, 1379-1998, 1387-1618, 1391-1956, 1399-1822, 1404-1679, 1409-1993, 1413-1987, 1415-1677, 1419-1704, 1467-1673, 1468-1792, 1473-1687, 1545-1787, 1559-1833, 1566-2120, 1582-2188, 1672-2009, 1686-1961, 1689-2385, 1692-2276, 1694-1926, 1694-2197, 1701-2085, 1717-1983, 1728-2315, 1728-2346, 1739-2362, 1740-2009, 1755-2365, 1757-2346, 1813-2450, 1825-2292, 1827-2390, 1828-2429, 1836-2445, 1838-2343, 1879-2293, 1902-2429, 1908-2554, 1913-2131, 1916-2185, 1920-2150, 1920-2159, 1921-2530, 1922-2148, 1958-2621, 1960-2224, 2007-2619, 2034-2553, 2065-2399, 2067-2657, 2085-2644, 2089-2264, 2089-2289, 2112-2397, 2115-2732, 2123-2677, 2125-2385, 2133-2395, 2133-2399,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
26/ 1760208CB1/ 3141 (continued)	2164-2415, 2177-2793, 2179-2369, 2184-2784, 2186-2707, 2188-2834, 2194-2771, 2199-2481, 2203-2768, 2207-2737, 2210-2580, 2217-2499, 2219-2754, 2232-2539, 2235-2490, 2235-2504, 2258-2748, 2272-2834, 2283-2833, 2290-2559, 2327-2607, 2341-2826, 2361-2957, 2392-2664, 2401-2641, 2404-2613, 2410-2795, 2416-3018, 2428-2656, 2439-3112, 2449-2765, 2462-2741, 2474-2709, 2500-3115, 2505-2855, 2509-2784, 2529-2820, 2532-2766, 2542-3116, 2554-3105, 2564-2819, 2575-2821, 2582-2821, 2584-2847, 2626-2830, 2626-2901, 2696-3102, 2712-3130, 2816-3063, 2842-2994, 2842-3061, 2952-3141, 2963-3141, 2987-3141
27/ 1900132CB1/ 3261	1-545, 139-345, 242-758, 242-911, 270-504, 323-373, 429-993, 465-986, 576-1153, 624-1183, 829-1452, 1025-1211, 1025-1512, 1215-1758, 1356-1619, 1356-1675, 1356-1815, 1356-1818, 1356-1820, 1356-1826, 1356-1831, 1356-1858, 1403-1652, 1423-1710, 1451-1882, 1463-1684, 1463-1697, 1463-1953, 1490-1945, 1513-1742, 1513-1766, 1544-1796, 1582-1811, 1605-1852, 1699-2201, 1759-2203, 1871-2129, 1904-2245, 1936-2354, 1976-2235, 1982-2204, 1982-2493, 2050-2290, 2059-2367, 2124-2422, 2147-2521, 2147-2614, 2222-2460, 2228-2673, 2253-2545, 2273-2866, 2290-2535, 2290-2770, 2351-2555, 2446-2912, 2448-2731, 2579-2974, 2579-3223, 2591-2841, 2608-2844, 2608-2846, 2632-2894, 2652-3206, 2700-3231, 2715-3229, 2732-3000, 2738-2953, 2740-3224, 2790-2990, 2793-3261, 2794-3253, 2804-3093, 2811-3255, 2833-3071, 2836-3143, 2865-3102, 2884-3227, 2891-3088, 2913-3154, 2990-3234, 3063-3256
28/7487551CB1/ 1097	1-735, 120-770, 215-770, 606-1097
29/ 1871014CB1/ 1633	1-265, 44-231, 44-304, 44-313, 47-295, 50-677, 51-318, 52-506, 56-253, 56-273, 56-287, 56-292, 56-293, 56-307, 56-316, 56-317, 56-323, 57-292, 57-331, 58-297, 59-289, 59-332, 59-513, 59-518, 59-553, 60-340, 60-345, 60-351, 62-352, 62-509, 64-312, 64-357, 65-316, 65-323, 65-330, 65-336, 65-354, 65-360, 65-365, 65-450, 65-655, 66-368, 66-533, 68-228, 68-256, 69-324, 69-329, 69-374, 69-537, 69-540, 70-303, 70-318, 71-341, 71-361, 74-655, 77-225, 77-309, 77-632, 78-340, 78-368, 79-310, 80-388, 80-553, 86-362, 89-360, 92-345, 102-357, 106-425, 114-471, 121-329, 128-451, 156-420, 165-665, 174-806, 175-806, 193-403, 200-816, 235-472, 237-516, 243-477, 285-742, 317-569, 333-617, 335-888, 342-853, 353-964, 362-517, 372-655, 409-625, 410-593, 410-1104, 418-660, 419-582, 419-940, 425-675, 431-642, 431-1119, 450-777, 454-974, 461-703, 470-746, 486-733, 506-1064, 506-1125, 506-1178, 508-678, 509-934, 512-798, 524-875, 555-806, 563-833, 574-793, 575-798, 575-863,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
29/ 1871014CB1/ 1633 (continued)	575-1048, 576-877, 580-771, 580-837, 593-816, 595-1181, 601-1083, 603-853, 623-1185, 624-890, 626-884, 639-928, 662-938, 664-998, 693-969, 728-1005, 730-1281, 734-1310, 738-1347, 741-1075, 742-1039, 757-1034, 760-1064, 773-1338, 781-960, 783-1255, 784-1042, 784-1047, 786-1029, 788-1055, 788-1073, 803-1049, 803-1060, 806-1021, 820-1078, 826-1113, 826-1187, 826-1354, 848-1091, 861-1088, 892-1050, 917-1578, 929-1134, 947-1206, 991-1226, 998-1241, 1011-1609, 1052-1612, 1059-1322, 1068-1546, 1101-1559, 1117-1587, 1124-1587, 1125-1404, 1126-1571, 1273-1587, 1300-1526, 1349-1633, 1352-1594, 1420-1587
30/ 2903166CB1/ 5869	1-682, 218-450, 236-400, 237-305, 237-338, 237-345, 237-349, 237-355, 237-356, 237-357, 237-358, 237-360, 237-362, 237-365, 237-366, 237-369, 237-374, 237-378, 237-380, 237-381, 237-384, 237-386, 237-387, 237-388, 237-390, 237-391, 237-392, 237-398, 237-400, 237-403, 237-406, 237-413, 237-414, 237-416, 237-419, 237-421, 237-422, 237-425, 237-427, 237-429, 237-430, 237-436, 237-444, 237-447, 237-452, 237-472, 237-482, 237-486, 237-490, 237-496, 237-503, 237-617, 237-622, 237-623, 237-651, 237-699, 237-707, 237-746, 238-5669, 239-513, 239-648, 249-496, 267-522, 271-491, 280-537, 311-400, 328-510, 334-905, 342-489, 384-619, 395-511, 407-702, 411-900, 411-905, 445-735, 447-678, 448-687, 448-725, 475-734, 499-604, 519-828, 527-833, 535-822, 566-842, 583-779, 584-831, 593-905, 595-905, 599-905, 602-857, 605-867, 627-905, 644-889, 649-794, 649-879, 649-905, 659-859, 662-905, 678-905, 683-905, 686-905, 692-905, 695-791, 696-893, 717-905, 723-790, 723-798, 741-905, 1040-1263, 1040-1270, 1040-1295, 1040-1296, 1040-1306, 1040-1310, 1040-1311, 1040-1315, 1049-1213, 1073-1699, 1116-1290, 1151-1310, 1197-1310, 1466-1941, 1468-1635, 1696-2170, 1696-2321, 2372-2833, 2423-2860, 2458-2493, 2458-2979, 2468-2728, 2471-2899, 2516-3125, 2565-2809, 2604-2779, 2604-2797, 2607-2731, 2607-2782, 2607-2797, 2607-2821, 2607-2899, 2607-2957, 2607-3179, 2625-2711, 2625-2747, 2625-2857, 2625-2875, 2633-2875, 2643-3084, 2664-2719, 2667-2719, 2667-2743, 2667-2755, 2667-2797, 2667-2821, 2667-2875, 2676-2719, 2676-2756, 2676-2875, 2685-2806, 2685-2887, 2685-2911, 2685-2938, 2685-2977, 2685-3035, 2685-3287, 2700-3179, 2704-2788, 2706-2735, 2706-2751, 2706-2825, 2707-2788, 2713-2796, 2713-2797, 2713-2953, 2714-2797, 2719-2834, 2721-2953, 2745-2953, 2745-2977, 2745-3035, 2777-3314, 2782-2866, 2784-2813, 2784-2829, 2784-2903, 2784-3047, 2785-2866, 2791-3047, 2792-2875, 2792-2957, 2797-2912, 2799-3047, 2823-2884, 2823-2965, 2823-2989, 2823-3014, 2823-3031, 2855-3287, 2862-2891, 2862-2907, 2862-2944, 2862-2981, 2862-3031, 2863-2944,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
30/ 2903166CB1/ 5869 (continued)	2869-2899, 2869-2957, 2869-3031, 2870-2957, 2870-3558, 2875-2990, 2877-3031, 2901-2960, 2901-3031, 2938-3022, 2940-3032, 2941-3022, 2947-3032, 2956-3032, 2976-3031, 2979-3031, 2994-3579, 3021-3062, 3021-3115, 3023-3062, 3028-3169, 3030-3116, 3033-3062, 3040-3139, 3040-3179, 3055-3248, 3057-3611, 3070-3169, 3077-3161, 3094-3125, 3094-3169, 3094-3182, 3094-3186, 3094-3200, 3094-3277, 3099-3151, 3102-3176, 3106-3171, 3148-3233, 3148-3285, 3153-3287, 3163-3287, 3165-3287, 3177-3207, 3178-3277, 3180-3225, 3180-3278, 3185-3364, 3192-3286, 3192-3373, 3192-3396, 3192-3424, 3192-3442, 3195-3233, 3195-3268, 3195-3286, 3195-3287, 3195-3310, 3195-3312, 3195-3346, 3195-3364, 3195-3382, 3195-3388, 3195-3389, 3195-3442, 3195-3458, 3199-3323, 3207-3650, 3210-3322, 3214-3364, 3243-3442, 3249-3373, 3249-3424, 3249-3442, 3263-3302, 3263-3310, 3263-3364, 3263-3388, 3263-3442, 3268-3355, 3271-3442, 3280-3363, 3280-3442, 3292-3364, 3292-3442, 3297-3388, 3298-3364, 3309-3442, 3312-3442, 3344-3442, 3349-3424, 3351-3403, 3351-3423, 3352-3424, 3358-3389, 3358-3443, 3358-3541, 3359-3511, 3366-3425, 3366-3443, 3399-3442, 3400-3497, 3428-3541, 3432-3541, 3442-3711, 3442-3774, 3454-3496, 3454-3541, 3457-3504, 3457-3532, 3457-3541, 3459-3497, 3459-3541, 3460-3540, 3466-3541, 3473-3532, 3474-3535, 3474-3541, 3533-3753, 3556-3721, 3616-3978, 3732-4231, 3978-4520, 4068-4652, 4085-4617, 4112-4399, 4112-4625, 4188-4469, 4206-4695, 4241-4404, 4338-4807, 4381-4979, 4390-4979, 4463-4714, 4463-4720, 4463-4961, 4487-4722, 4504-4980, 4536-4778, 4536-4963, 4562-4844, 4565-5199, 4629-4890, 4631-4934, 4650-4856, 4822-4857, 4877-5139, 4916-5483, 4929-5179, 5004-5260, 5059-5811, 5062-5285, 5070-5346, 5105-5390, 5107-5328, 5120-5337, 5151-5647, 5207-5509, 5207-5627, 5209-5516, 5214-5796, 5247-5529, 5258-5833, 5302-5834, 5312-5821, 5323-5834, 5325-5578, 5333-5393, 5337-5865, 5350-5858, 5358-5861, 5366-5738, 5366-5869, 5371-5598, 5371-5671, 5371-5835, 5372-5648, 5374-5851, 5379-5864, 5393-5844, 5408-5439, 5416-5865, 5418-5858, 5454-5869, 5467-5795

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
31/ 1723804CB1/ 3879	1-573, 8-202, 43-556, 44-633, 59-598, 91-556, 203-384, 203-699, 255-829, 304-562, 382-924, 690-962, 690-1148, 904-1198, 938-1198, 961-1195, 961-1239, 1075-1643, 1165-1765, 1182-1761, 1321-1570, 1323-1781, 1361-1807, 1377-1781, 1432-1781, 1442-1963, 1443-1782, 1496-1781, 1525-1781, 1682-1954, 1682-2025, 1682-2292, 1759-2216, 1790-2377, 1820-2346, 1850-2233, 1908-2184, 1908-2453, 1931-2549, 1956-2331, 1970-2271, 1970-2361, 1970-2414, 1970-2435, 1970-2456, 1970-2462, 1970-2463, 1970-2499, 1970-2527, 1970-2545, 1978-2565, 1989-2238, 2004-2549, 2026-2463, 2026-2470, 2026-2544, 2061-2549, 2078-2367, 2080-2356, 2145-2763, 2151-2549, 2176-2405, 2195-2769, 2276-2671, 2302-2906, 2322-2841, 2322-2857, 2324-2577, 2324-2706, 2327-2772, 2330-2578, 2330-2580, 2330-2606, 2330-2640, 2331-2529, 2331-2605, 2331-2640, 2333-2559, 2333-2636, 2337-2622, 2337-2698, 2337-2790, 2337-2945, 2338-2848, 2339-2591, 2340-2652, 2342-2599, 2344-2781, 2370-2573, 2398-2620, 2398-2658, 2398-2825, 2398-2855, 2398-2869, 2398-2905, 2398-2913, 2444-2600, 2445-2771, 2445-2839, 2459-2741, 2467-2715, 2479-2758, 2480-2700, 2480-2785, 2492-2746, 2504-2807, 2504-3042, 2505-2729, 2523-2763, 2544-3123, 2548-2833, 2549-2827, 2550-2793, 2558-2841, 2582-2802, 2603-2868, 2603-3168, 2629-2818, 2661-2885, 2696-2943, 2722-2935, 2761-2930, 2789-3102, 2793-3026, 2827-3074, 2861-3093, 2861-3096, 2861-3099, 2861-3108, 2861-3110, 2861-3111, 2861-3142, 2901-3177, 2901-3191, 2905-3203, 2921-3127, 2921-3554, 2942-3168, 2971-3398, 2978-3538, 2981-3243, 2992-3162, 2992-3403, 3016-3310, 3016-3315, 3027-3258, 3075-3351, 3084-3341, 3121-3413, 3132-3357, 3134-3337, 3231-3859, 3243-3842, 3365-3600, 3371-3606, 3371-3812, 3371-3850, 3372-3848, 3385-3629, 3395-3608, 3411-3630, 3412-3625, 3412-3879, 3456-3708, 3567-3778, 3632-3853, 3657-3839
32/ 7736769CB1/ 2160	1-160, 1-1764, 42-365, 52-290, 73-278, 110-760, 194-772, 243-813, 364-963, 463-982, 481-923, 499-1048, 592-1201, 658-985, 658-1026, 658-1055, 658-1084, 658-1107, 658-1117, 658-1122, 658-1148, 658-1154, 658-1167, 658-1201, 660-1079, 666-1041, 666-1047, 666-1191, 676-1049, 676-1068, 681-1067, 681-1152, 682-1132, 715-1202, 762-991, 809-1045, 833-1392, 837-1117, 850-1111, 872-1124, 872-1128, 873-1136, 880-1087, 891-1152, 901-1162, 913-1507, 917-1193, 917-1212, 1058-1536, 1064-1262, 1067-1705, 1075-1360, 1076-1374, 1082-1725, 1088-1286, 1088-1383, 1099-1295, 1101-1367, 1101-1783, 1105-1252, 1113-1396, 1113-1573, 1132-1409, 1155-1396, 1198-1448, 1209-1463, 1220-1430, 1249-1518, 1266-1528, 1266-1766, 1330-1543, 1334-1578, 1334-1906, 1335-1582, 1355-1600, 1358-1586, 1380-1508, 1393-1420,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
32/ 7736769CB1/ 2160 (continued)	1443-1570, 1486-1756, 1490-1729, 1493-2120, 1494-1743, 1539-1776, 1547-1772, 1551-1783, 1562-1774, 1562-1809, 1577-1817, 1606-1834, 1650-1876, 1662-1877, 1664-1912, 1701-1967, 1743-1950, 1746-2160, 1779-1992, 1779-2022, 1797-1999, 1811-2033, 1894-2160, 1958-2160, 1963-2160, 1968-2160, 1970-2160, 1983-2160, 1995-2160, 2004-2160, 2015-2160
33/ 7492451CB1/ 2800	1-36, 1-640, 40-85, 40-125, 40-132, 40-133, 40-134, 54-134, 61-103, 64-134, 87-134, 152-680, 152-757, 164-625, 165-349, 172-577, 184-759, 222-821, 547-585, 547-849, 549-1045, 560-585, 632-657, 632-977, 669-941, 718-1327, 722-1159, 728-1263, 749-995, 749-1127, 770-1283, 794-1322, 851-1112, 874-1162, 895-1152, 950-1230, 971-1280, 991-1336, 992-1224, 993-1260, 993-1316, 1023-1501, 1036-1473, 1045-1558, 1049-1303, 1049-1675, 1065-1161, 1079-1365, 1087-1626, 1090-1384, 1090-1423, 1114-1540, 1189-1384, 1204-1352, 1219-1500, 1231-1703, 1241-1856, 1244-1914, 1246-1817, 1248-1381, 1248-1685, 1252-1363, 1306-1569, 1332-1906, 1338-1599, 1338-1625, 1338-1636, 1346-1570, 1355-1961, 1362-1426, 1362-1646, 1362-1875, 1372-1653, 1383-1656, 1390-1641, 1390-1781, 1390-1934, 1393-1672, 1408-1625, 1412-1670, 1423-1683, 1423-1686, 1432-1683, 1437-1684, 1443-1718, 1447-1934, 1476-1584, 1480-2020, 1491-2034, 1493-1966, 1505-1760, 1505-1820, 1513-2083, 1513-2122, 1520-1757, 1530-2016, 1533-1831, 1535-1941, 1537-1557, 1543-1867, 1548-1608, 1548-1610, 1557-1593, 1566-2152, 1568-1870, 1569-1716, 1569-2051, 1572-1939, 1580-2157, 1608-2180, 1616-1878, 1637-2160, 1639-1924, 1639-2287, 1648-1896, 1657-1917, 1657-2227, 1658-1932, 1658-2240, 1660-1929, 1661-1960, 1662-1909, 1672-1762, 1672-1813, 1675-2021, 1682-2087, 1686-1952, 1686-1995, 1686-2268, 1690-1940, 1690-1948, 1696-2188, 1705-1851, 1705-2333, 1719-1942, 1720-1959, 1723-2251, 1724-2275, 1726-1950, 1726-1990, 1739-1813, 1748-1973, 1748-2033, 1748-2034, 1750-2227, 1750-2424, 1759-2321, 1782-2363, 1783-1972, 1790-1926, 1794-1899, 1800-2087, 1800-2400, 1801-2054, 1841-2097, 1883-2228, 1910-2164, 1910-2495, 1928-2192, 1932-1966, 1935-1966, 1940-2448, 1950-2181, 1950-2446, 1950-2478, 1962-2206, 1987-2018, 2013-2463, 2024-2198, 2026-2180, 2027-2137, 2029-2744, 2034-2535, 2044-2628, 2058-2719, 2076-2740, 2079-2782, 2096-2724, 2114-2692, 2126-2164, 2129-2160, 2129-2164, 2143-2623, 2145-2223, 2151-2742, 2182-2792, 2183-2800, 2189-2669, 2191-2459, 2193-2450, 2201-2766, 2203-2496, 2205-2448, 2205-2486, 2206-2465, 2206-2783, 2206-2792, 2212-2477, 2214-2498, 2231-2514, 2231-2800, 2234-2493, 2235-2476, 2235-2535, 2235-2764, 2235-2781, 2264-2538, 2264-2542, 2277-2534, 2284-2710, 2284-2711, 2284-2800, 2290-2547, 2290-2552, 2302-2333,

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
33/7492451CB1/ 2800 (continued)	2302-2337, 2302-2340, 2307-2733, 2308-2533, 2311-2763, 2315-2528, 2315-2717, 2320-2769, 2321-2399, 2322-2539, 2322-2541, 2323-2545, 2329-2578, 2329-2594, 2329-2626, 2331-2551, 2332-2798, 2333-2794, 2340-2437, 2341-2432, 2341-2444, 2341-2451, 2341-2467, 2344-2533, 2344-2749, 2344-2763, 2344-2796, 2344-2800, 2345-2797, 2346-2641, 2351-2605, 2351-2635, 2356-2795, 2357-2470, 2363-2616, 2363-2800, 2373-2624, 2376-2800, 2380-2695, 2382-2675, 2384-2585, 2385-2791, 2386-2791, 2387-2792, 2388-2797, 2390-2794, 2392-2651, 2393-2759, 2393-2787, 2406-2797, 2410-2650, 2410-2677, 2410-2679, 2410-2684, 2413-2640, 2559-2599, 2732-2774
34/ 4650669CB1/ 1384	1-166, 1-170, 84-346, 84-353, 103-702, 194-472, 280-454, 280-477, 280-708, 305-574, 406-666, 406-740, 406-915, 406-972, 438-1098, 487-766, 509-1058, 536-1098, 545-980, 601-916, 626-886, 674-1073, 688-852, 729-1228, 737-1299, 756-1246, 839-1353, 918-1384, 942-1374, 959-1377, 960-1365, 961-1361, 966-1368, 976-1371, 976-1384, 977-1371, 992-1364, 1002-1380, 1045-1383
35/ 7485268CB1/ 969	1-391, 211-969
36/ 2112995CB1/ 2792	1-458, 169-654, 171-688, 171-972, 172-594, 172-616, 172-748, 173-332, 173-338, 173-661, 173-769, 174-485, 226-746, 310-508, 394-934, 436-692, 512-879, 585-842, 585-1393, 615-1184, 617-785, 666-1369, 675-859, 691-1219, 711-1336, 824-1410, 824-1434, 825-1475, 836-1462, 864-1485, 867-1576, 942-1505, 943-1674, 978-1514, 986-1491, 1024-1743, 1058-1604, 1076-1644, 1146-1708, 1215-1491, 1219-1797, 1288-1919, 1314-1793, 1326-1841, 1352-1596, 1352-1788, 1442-1916, 1455-1706, 1469-1754, 1488-1993, 1489-1611, 1493-1844, 1528-1786, 1643-1838, 1680-1928, 1691-2233, 1692-2091, 1731-2407, 1780-2329, 1888-2503, 1924-2524, 1935-2167, 1935-2367, 1935-2432, 1955-2203, 2066-2354, 2066-2367, 2109-2336, 2119-2398, 2151-2770, 2155-2787, 2160-2377, 2162-2428, 2162-2443, 2253-2503, 2260-2785, 2276-2543, 2276-2570, 2276-2572, 2276-2580, 2276-2584, 2276-2587, 2276-2593, 2277-2595, 2278-2582, 2278-2593, 2286-2561, 2377-2628, 2386-2643, 2386-2690, 2399-2790, 2481-2792

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
37/ 1613452CB1/ 3567	1-259, 1-321, 1-639, 8-259, 38-289, 64-316, 69-370, 201-2103, 287-436, 343-811, 853-1095, 880-1414, 909-1301, 923-1286, 924-1205, 1106-1338, 1211-1440, 1211-1772, 1317-1612, 1341-1580, 1370-1638, 1460-1685, 1587-1806, 1613-1776, 1613-1965, 1702-1928, 1741-1968, 1741-2264, 1790-2065, 1856-2142, 1890-2343, 1944-2225, 1987-2246, 2001-2305, 2045-2318, 2048-2304, 2056-2431, 2073-2167, 2171-2430, 2193-2442, 2196-2516, 2231-2505, 2263-2528, 2263-2652, 2306-2560, 2315-2500, 2315-2845, 2318-2444, 2318-2564, 2318-2872, 2339-2787, 2339-2802, 2339-2820, 2410-2587, 2412-2681, 2412-2906, 2442-2903, 2484-2760, 2491-2677, 2536-2808, 2547-2864, 2635-2761, 2635-2862, 2655-2811, 2659-2765, 2659-2940, 2663-2947, 2669-3232, 2673-2933, 2673-3370, 2681-3210, 2684-2948, 2691-2919, 2691-3218, 2728-3009, 2741-3036, 2755-3410, 2756-3010, 2756-3040, 2757-3289, 2779-3327, 2782-3433, 2789-3433, 2805-3079, 2827-3422, 2850-3140, 2863-3105, 2864-3167, 2875-3134, 2880-3380, 2882-3418, 2898-3125, 2906-3433, 2912-3551, 2926-3422, 2966-3407, 2976-3415, 2977-3092, 2984-3433, 2995-3389, 2995-3412, 3013-3430, 3015-3410, 3016-3343, 3016-3383, 3016-3396, 3027-3235, 3039-3430, 3070-3558, 3076-3430, 3079-3422, 3079-3433, 3083-3300, 3093-3426, 3096-3550, 3101-3422, 3122-3398, 3130-3567, 3155-3433, 3155-3550, 3155-3560, 3176-3558, 3192-3563, 3225-3433, 3235-3421, 3263-3387, 3275-3433, 3308-3563, 3367-3430, 3374-3433, 3461-3567
38/ 55061615CB1/ 6004	1-740, 227-740, 615-1164, 620-1045, 620-1219, 620-1299, 644-1300, 915-1134, 915-1411, 915-1670, 977-1300, 1101-6004, 1620-1869, 1624-2104, 2247-2989, 2789-3386, 2831-3269, 2831-3348, 2831-3411, 2841-2897, 2843-2897, 2898-3244, 2898-3351, 3259-3413, 3306-3413, 3539-4052, 3539-4055, 3541-4055, 3558-4237, 3558-4253, 3599-4023, 3687-4414, 3706-4055, 3871-4055, 3890-4055, 3941-4586, 3942-4586, 4012-4586, 4081-4542, 4279-4550, 4593-5209, 4596-5186, 4778-5310, 4780-5310, 4790-4913, 4790-5254, 4790-5299, 4790-5309, 4845-4913, 4860-4913, 4863-4913, 4937-5310, 4941-5309, 4942-5171, 4943-5310, 4949-5310, 4952-5310, 4955-5310, 4956-5310, 5018-5310, 5058-5310, 5078-5310, 5177-5310, 5261-5310, 5384-5425, 5384-5465, 5384-5485, 5384-5558, 5384-5560, 5386-5459, 5386-5556, 5423-5558, 5628-5718, 5631-5865, 5634-5831, 5634-5864, 5634-5865, 5634-5866, 5634-5885, 5637-5831, 5641-5865

Table 4 (cont.)

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
39/ 7503435CB1/ 1917	1-241, 1-247, 1-350, 1-422, 1-529, 1-632, 1-634, 1-651, 1-666, 1-667, 1-688, 1-689, 1-795, 1-876, 1-887, 1-895, 1-1455, 1-1460, 7-266, 11-666, 16-173, 18-523, 23-322, 54-928, 55-640, 64-710, 140-369, 256-830, 259-502, 293-981, 306-903, 324-607, 341-831, 345-886, 359-477, 359-944, 366-747, 374-874, 376-622, 385-691, 385-747, 387-930, 397-1041, 401-678, 422-1040, 437-831, 479-601, 487-719, 487-1041, 650-955, 696-1011, 707-1037, 721-902, 751-1185, 759-1036, 1039-1360, 1041-1425, 1113-1338, 1134-1425, 1165-1425, 1179-1425, 1228-1374, 1228-1458, 1228-1468, 1228-1481, 1228-1917, 1229-1479, 1256-1472, 1282-1468, 1314-1468
40/ 7504149CB1/ 1208	1-541, 1-1208, 3-590, 3-791, 31-287, 31-520, 39-310, 144-407, 182-467, 209-1111, 226-1115, 239-499, 239-573, 239-748, 239-805, 248-1115, 248-1116, 270-401, 271-926, 280-1116, 300-1115, 320-377, 320-430, 320-599, 320-848, 320-880, 320-953, 342-891, 366-1106, 370-931, 378-813, 400-1115, 403-1111, 434-749, 439-545, 450-810, 452-916, 459-719, 507-906, 511-1083, 562-1060, 590-1080, 686-1208, 702-1165, 703-1032, 704-1165, 727-1121, 748-997, 751-1208, 762-1169, 775-1208, 792-1208, 793-1199, 794-1195, 799-1202, 809-1205, 810-1205, 825-1198, 835-1208, 847-1084, 862-1208, 878-1208, 893-1155, 1076-1202, 1093-1208, 1096-1206, 1096-1207

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
21	1419725CB1	KIDNNOT09
22	628613CB1	SINTFEE01
23	7111920CB1	BRSTNOT05
24	3072268CB1	SPLNFET02
25	5519523CB1	KERANOT01
26	1760208CB1	URETTUT01
27	1900132CB1	ISLTNOT01
28	7487551CB1	SINIDME01
29	1871014CB1	BRSTTUT02
30	2903166CB1	DRGCNOT01
31	1723804CB1	KERANOT01
32	7736769CB1	THP1NOB01
33	7492451CB1	LIVRNON08
34	4650669CB1	PROSTUT20
36	2112995CB1	PROSTUS23
37	1613452CB1	PROSNON01
38	55061615CB1	BRAIFER06
39	7503435CB1	KIDNNOT09
40	7504149CB1	BRONNOT02

Table 6

Library	Vector	Library Description
BRAIFER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRONNOT02	pINCY	Library was constructed using RNA isolated from right lower lobe bronchial tissue removed from a pool of 9 nonasthmatic Caucasian male and female donors, 18- to 55-years-old during bronchial pinch biopsies. Patient history included atopy as determined by positive skin tests to common aero-allergens with no bronchial hyperresponsiveness to histamine. The donors were not current smokers and had no history of alcohol or drug abuse.
BRSTNOT05	PSPORT1	Library was constructed using RNA isolated from breast tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated multicentric invasive grade 4 lobular carcinoma. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular and cardiovascular disease, breast and prostate cancer, and type I diabetes.
BRSTTUT02	PSPORT1	Library was constructed using RNA isolated from breast tumor tissue removed from a 54-year-old Caucasian female during a bilateral radical mastectomy with reconstruction. Pathology indicated residual invasive grade 3 mammary ductal adenocarcinoma. The remaining breast parenchyma exhibited proliferative fibrocystic changes without atypia. One of 10 axillary lymph nodes had metastatic tumor as a microscopic intranodal focus. Patient history included kidney infection and condyloma acuminatum. Family history included benign hypertension, hyperlipidemia, and a malignant colon neoplasm.
DRGCNOT01	pINCY	Library was constructed using RNA isolated from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.
ISLTNOT01	pINCY	Library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.

Table 6 (cont.)

Library	Vector	Library Description
KERANOT01	PBLUESCRIPT	Library was constructed using RNA isolated from neonatal keratinocytes obtained from the leg skin of a spontaneously aborted black male.
KIDNNOT09	pINCY	Library was constructed using RNA isolated from the kidney tissue of a Caucasian male fetus, who died at 23 weeks' gestation.
LIVRNON08	pINCY	This normalized library was constructed from 5.7 million independent clones from a pooled liver tissue library. Starting RNA was made from pooled liver tissue removed from a 4-year-old Hispanic male who died from anoxia and a 16 week female fetus who died after 16-weeks gestation from anencephaly. Serologies were positive for cytolomegalovirus in the 4-year-old. Patient history included asthma in the 4-year-old. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother of the fetus. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
PROSNON01	PSPORT1	This normalized prostate library was constructed from 4.4 M independent clones from a prostate library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.
PROSTUS23	pINCY	This subtracted prostate tumor library was constructed using 10 million clones from a pooled prostate tumor library that was subjected to 2 rounds of subtractive hybridization with 10 million clones from a pooled prostate tissue library. The starting library for subtraction was constructed by pooling equal numbers of clones from 4 prostate tumor libraries using mRNA isolated from prostate tumor removed from Caucasian males at ages 58 (A), 61 (B), 66 (C), and 68 (D) during prostatectomy with lymph node excision. Pathology indicated adenocarcinoma in all donors. History included elevated PSA, induration and tobacco abuse in donor A; elevated PSA, induration, prostate hyperplasia, renal failure, osteoarthritis, renal artery stenosis, benign HTN, thrombocytopenia, hyperlipidemia, tobacco/alcohol abuse and hepatitis C (carrier) in donor B; elevated PSA, induration, and tobacco abuse in donor C; and elevated PSA, induration, hypercholesterolemia, and kidney calculus in donor D. The hybridization probe for

Table 6 (cont.)

Library	Vector	Library Description
		subtraction was constructed by pooling equal numbers of cDNA clones from 3 prostate tissue libraries derived from prostate tissue, prostate epithelial cells, and fibroblasts from prostate stroma from 3 different donors. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR 19 (1991):1954 and Bonaldo, et al. Genome Research 6 (1996):791.
PROSTUT20	pINCY	The library was constructed using RNA isolated from prostatetumor tissue removed from a 58-year-old Caucasian male during radical prostatectomy,regional lymph node excision, and prostate needle biopsy. Pathology indicatedadenocarcinoma (Gleason grade 3+2) of the prostate, which formed a predominant massinvolving primarily the right side and focally involved the left side, peripherallyand anteriorly. The patient presented with elevated prostate specific antigen (PSA)and induration. Family history included breast cancer .
SINIDME01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from diseased ileum tissue removed from a 29-year-old Caucasian female during jejunostomy. Pathology indicated mild chronic inflammation. The patient presented with ulcerative colitis. Patient history included a benign neoplasm of the large bowel. Patient medications included Asacol, Rowasa, Clomid and Pergonol. Family history included benign hypertension in the mother, and colon cancer and cerebrovascular accident in the grandparent(s).
SINTFEE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from fetal demise.
SPLNFET02	pINCY	Library was constructed using RNA isolated from spleen tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.
THP1NOB01	PBLUESCRIPT	"Library was constructed using RNA isolated from cultured, unstimulated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer (1980) 26:171)."
URETTUT01	pINCY	Library was constructed using RNA isolated from right ureter tumor tissue of a 69-year-old Caucasian male during ureterectomy and lymph node excision. Pathology indicated invasive grade 3 transitional cell carcinoma. Patient history included benign colon neoplasm, tobacco use, asthma, emphysema, acute duodenal ulcer, and hyperplasia of the prostate. Family history included atherosclerotic coronary artery disease, congestive heart failure, and malignant lung neoplasm.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant

polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

19. A method for treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional MDDT, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of MDDT in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of MDDT in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of MDDT in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides

of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

<110> INCYTE GENOMICS, INC.
LAL, Preeti G.
BAUGHN, Mariah R.
YAO, Monique G.
WALIA, Narinder K.
ELLIOTT, Vicki S.
XU, Yuming
HONCHELL, Cynthia D.
YUE, Henry
DING, Li
GIETZEN, Kimberly J.
ISON, Craig H.
LU, Dyung Aina M.
HAFALIA, April J.A.
GANDHI, Ameena R.
THANGAVELU, Kavitha
SANJANWALA, Madhusudan M.
TANG, Y. Tom
RAMKUMAR, Jayalaxmi
GRIFFIN, Jennifer A.
SWARNAKER, Anita
AZIMZAI, Yalda
SAPPERSTEIN, Stephanie K.
BURFORD, Neil
LEE, Ernestine A.
LU, Yan
TRAN, Uyen K.
MARQUIS, Joseph P.

<120> MOLECULES FOR DISEASE DETECTION AND TREATMENT

<130> PF-0899 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/268,117; 60/269,618; 60/271,118; 60/274,486;
60/274,436; 60/334,229; UnAssigned
<151> 2001-02-09; 2001-02-15; 2001-02-23; 2001-03-07
2001-03-09; 2001-11-28; 2002-02-01

<160> 40
<170> PERL Program

<210> 1
<211> 198
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1419725CD1

<400> 1
Met Lys Ser Lys Gly Val Lys Ser Tyr Gln Arg Arg Pro Arg Glu
1 5 10 15

```
Glu Arg Thr Gln Arg Arg Thr Arg Cys Gln Ser Arg Arg Gly Ser
                  20                  25                  30
Trp Arg Ser Arg His Trp Arg Trp Trp Asn Lys Leu Leu Pro Thr
                  35                  40                  45
Pro Trp Met Thr Gly Thr Leu Gly Ser Ser Ser Cys Gln Ala Ser
                  50                  55                  60
Leu Ala Met Cys Pro Ala Pro Ala Ser Ser Ser Ala Pro Ala Phe
                  65                  70                  75
Leu Cys Ser Pro Thr Arg His Cys Arg Asn Leu Gly Arg Ser Thr
                  80                  85                  90
His Gln Ala Val Pro Arg Thr Pro Asn Ile Ser Pro His Phe Pro
                  95                 100                 105
Glu His Thr Leu Arg Thr Trp Val Phe Tyr Leu Thr Met Gly Ala
                 110                 115                 120
Thr Cys Gln Gly Ile Ser Ser Ser Leu Ala Thr His Leu Ala Ile
                 125                 130                 135
Ser Pro Met Met Leu Trp Ala Ser Ala Pro Ser Arg Ser Ser Ser
                 140                 145                 150
Trp Leu Arg Pro Leu Asp Ile Lys Phe Pro Ser Leu Phe Ile Leu
                 155                 160                 165
Ser Gln Pro Ser Phe Trp Lys Gly Glu Arg Trp Val Gly Gly Trp
                 170                 175                 180
Glu Gly Gly Gly Thr Gln Arg Glu Asn Gly Leu Glu Ala Glu His
                 185                 190                 195
Leu Phe Tyr
```

<210> 2
<211> 385
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 628613CD1

<400> 2

```
Met Ser Leu Ile Leu Asn Ile Leu Arg Glu Met Leu Glu Tyr Phe
  1               5                  10                  15
Gly Val Pro Val Glu Gln Val Leu Leu Ile Trp Glu Asn Lys Asp
                  20                  25                  30
Tyr Gly Ser Thr Arg Ser Ile Val Arg Ile Ile Gly Lys Met Leu
                  35                  40                  45
Pro Leu Glu Pro Cys Arg Arg Pro Asn Phe Glu Leu Ile Pro Leu
                  50                  55                  60
Leu Asn Ser Val Asp Ser Asp Asn Cys Gly Ser Met Val Pro Ser
                  65                  70                  75
Phe Ala Asp Ile Leu Tyr Val Ala Asn Asp Glu Glu Ala Ser Tyr
                  80                  85                  90
Leu Arg Phe Arg Asn Ser Ile Trp Lys Asn Glu Glu Glu Lys Val
                  95                 100                 105
Glu Ile Phe His Pro Leu Arg Leu Val Arg Asp Pro Leu Ser Pro
                 110                 115                 120
Ala Val Arg Gln Lys Glu Thr Val Lys Asn Asp Leu Pro Val Asn
                 125                 130                 135
Glu Ala Ala Ile Arg Lys Ile Ala Ala Leu Glu Asn Glu Leu Thr
```

```
                140                 145                 150
Phe Leu Arg Ser Gln Ile Ala Ala Ile Val Glu Met Gln Glu Leu
                155                 160                 165
Lys Asn Ser Thr Asn Ser Ser Ser Phe Gly Leu Ser Asp Glu Arg
                170                 175                 180
Ile Ser Leu Gly Gln Leu Ser Ser Ser Arg Ala Ala His Leu Ser
                185                 190                 195
Val Asp Pro Asp Gln Leu Pro Gly Ser Val Leu Ser Pro Pro Pro
                200                 205                 210
Pro Pro Pro Leu Pro Pro Gln Phe Ser Ser Leu Gln Pro Pro Cys
                215                 220                 225
Phe Pro Pro Val Gln Pro Gly Ser Asn Asn Ile Cys Asp Ser Asp
                230                 235                 240
Asn Pro Ala Thr Glu Met Ser Lys Gln Asn Pro Ala Ala Asn Lys
                245                 250                 255
Thr Asn Tyr Ser His His Ser Lys Ser Gln Arg Asn Lys Asp Ile
                260                 265                 270
Pro Asn Met Leu Asp Val Leu Lys Asp Met Asn Lys Val Lys Leu
                275                 280                 285
Arg Ala Ile Glu Arg Ser Pro Gly Gly Arg Pro Ile His Lys Arg
                290                 295                 300
Lys Arg Gln Asn Ser His Trp Asp Pro Val Ser Leu Ile Ser His
                305                 310                 315
Ala Leu Lys Gln Lys Phe Ala Phe Gln Glu Asp Asp Ser Phe Glu
                320                 325                 330
Lys Glu Asn Arg Ser Trp Glu Ser Ser Pro Phe Ser Ser Pro Glu
                335                 340                 345
Thr Ser Arg Phe Gly His His Ile Ser Gln Ser Glu Gly Gln Arg
                350                 355                 360
Thr Lys Glu Glu Met Val Asn Thr Lys Ala Val Asp Gln Gly Ile
                365                 370                 375
Ser Asn Thr Ser Leu Leu Asn Ser Arg Ile
                380                 385
```

<210> 3
<211> 725
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7111920CD1

<400> 3

```
Met Val Met Lys Ala Ser Val Asp Asp Asp Asp Ser Gly Trp Glu
  1               5                  10                  15
Leu Ser Met Pro Glu Lys Met Glu Lys Ser Asn Thr Asn Trp Val
                 20                  25                  30
Asp Ile Thr Gln Asp Phe Glu Glu Ala Cys Arg Glu Leu Lys Leu
                 35                  40                  45
Gly Glu Leu Leu His Asp Lys Leu Phe Gly Leu Phe Glu Ala Met
                 50                  55                  60
Ser Ala Ile Glu Met Met Asp Pro Lys Met Asp Ala Gly Met Ile
                 65                  70                  75
Gly Asn Gln Val Asn Arg Lys Val Leu Asn Phe Glu Gln Ala Ile
                 80                  85                  90
```

```
Lys Asp Gly Thr Ile Lys Ile Lys Asp Leu Thr Leu Pro Glu Leu
                 95                 100                 105
Ile Gly Ile Met Asp Thr Cys Phe Cys Cys Leu Ile Thr Trp Leu
            110                 115                 120
Glu Gly His Ser Leu Ala Gln Thr Val Phe Thr Cys Leu Tyr Ile
            125                 130                 135
His Asn Pro Asp Phe Ile Glu Asp Pro Ala Met Lys Ala Phe Ala
            140                 145                 150
Leu Gly Ile Leu Lys Ile Cys Asp Ile Ala Arg Glu Lys Val Asn
            155                 160                 165
Lys Ala Ala Val Phe Glu Glu Glu Asp Phe Gln Ser Met Thr Tyr
            170                 175                 180
Gly Phe Lys Met Ala Asn Ser Val Thr Asp Leu Arg Val Thr Gly
            185                 190                 195
Met Leu Lys Asp Val Glu Asp Asp Met Gln Arg Arg Val Lys Ser
            200                 205                 210
Thr Arg Ser Arg Gln Gly Glu Glu Arg Asp Pro Glu Val Glu Leu
            215                 220                 225
Glu His Gln Gln Cys Leu Ala Val Phe Ser Arg Val Lys Phe Thr
            230                 235                 240
Arg Val Leu Leu Thr Val Leu Ile Ala Phe Thr Lys Lys Glu Thr
            245                 250                 255
Ser Ala Val Ala Glu Ala Gln Lys Leu Met Val Gln Ala Ala Asp
            260                 265                 270
Leu Leu Ser Ala Ile His Asn Ser Leu His His Gly Ile Gln Ala
            275                 280                 285
Gln Asn Asp Thr Thr Lys Gly Asp His Pro Ile Met Met Gly Phe
            290                 295                 300
Glu Pro Leu Val Asn Gln Arg Leu Leu Pro Pro Thr Phe Pro Arg
            305                 310                 315
Tyr Ala Lys Ile Ile Lys Arg Glu Glu Met Val Asn Tyr Phe Ala
            320                 325                 330
Arg Leu Ile Asp Arg Ile Lys Thr Val Cys Glu Val Val Asn Leu
            335                 340                 345
Thr Asn Leu His Cys Ile Leu Asp Phe Phe Cys Glu Phe Ser Glu
            350                 355                 360
Gln Ser Pro Cys Val Leu Ser Arg Ser Leu Leu Gln Thr Thr Phe
            365                 370                 375
Leu Val Asp Asn Lys Lys Val Phe Gly Thr His Leu Met Gln Asp
            380                 385                 390
Met Val Lys Asp Ala Leu Arg Ser Phe Val Ser Pro Pro Val Leu
            395                 400                 405
Ser Pro Lys Cys Tyr Leu Tyr Asn Asn His Gln Ala Lys Asp Cys
            410                 415                 420
Ile Asp Ser Phe Val Thr His Cys Val Arg Pro Phe Cys Ser Leu
            425                 430                 435
Ile Gln Ile His Gly His Asn Arg Ala Arg Gln Arg Asp Lys Leu
            440                 445                 450
Gly His Ile Leu Glu Glu Phe Ala Thr Leu Gln Asp Glu Ala Glu
            455                 460                 465
Lys Val Asp Ala Ala Leu His Thr Met Leu Leu Lys Gln Glu Pro
            470                 475                 480
Gln Arg Gln His Leu Ala Cys Leu Gly Thr Trp Val Leu Tyr His
            485                 490                 495
Asn Leu Arg Ile Met Ile Gln Tyr Leu Leu Ser Gly Phe Glu Leu
            500                 505                 510
```

```
Glu Leu Tyr Ser Met His Glu Tyr Tyr Tyr Ile Tyr Trp Tyr Leu
                515                 520                 525
Ser Glu Phe Leu Tyr Ala Trp Leu Met Ser Thr Leu Ser Arg Ala
                530                 535                 540
Asp Gly Ser Gln Met Ala Glu Glu Arg Ile Met Glu Glu Gln Gln
                545                 550                 555
Lys Gly Arg Ser Ser Lys Lys Thr Lys Lys Lys Lys Lys Val Arg
                560                 565                 570
Pro Leu Ser Arg Glu Ile Thr Met Ser Gln Ala Tyr Gln Asn Met
                575                 580                 585
Cys Ala Gly Met Phe Lys Thr Met Val Ala Phe Asp Met Asp Gly
                590                 595                 600
Lys Val Arg Lys Pro Lys Phe Glu Leu Asp Ser Glu Gln Val Arg
                605                 610                 615
Tyr Glu His Arg Phe Ala Pro Phe Asn Ser Val Met Thr Pro Pro
                620                 625                 630
Pro Val His Tyr Leu Gln Phe Lys Glu Met Ser Asp Leu Asn Lys
                635                 640                 645
Tyr Ser Pro Pro Pro Gln Ser Pro Glu Leu Tyr Val Ala Ala Ser
                650                 655                 660
Lys His Phe Gln Gln Ala Lys Met Ile Leu Glu Asn Ile Pro Asn
                665                 670                 675
Pro Asp His Glu Val Asn Arg Ile Leu Lys Val Ala Lys Pro Asn
                680                 685                 690
Phe Val Val Met Lys Leu Leu Ala Gly Gly His Lys Lys Glu Ser
                695                 700                 705
Lys Val Pro Pro Glu Phe Asp Phe Ser Ala His Lys Tyr Phe Pro
                710                 715                 720
Val Val Lys Leu Val
                725
```

<210> 4
<211> 332
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3072268CD1

<400> 4

```
Met Ala Leu Leu Cys Tyr Asn Arg Gly Cys Gly Gln Arg Phe Asp
  1               5                  10                  15
Pro Glu Thr Asn Ser Asp Asp Ala Cys Thr Tyr His Pro Gly Val
                 20                  25                  30
Pro Val Phe His Asp Ala Leu Lys Gly Trp Ser Cys Cys Lys Arg
                 35                  40                  45
Arg Thr Thr Asp Phe Ser Asp Phe Leu Ser Ile Val Gly Cys Thr
                 50                  55                  60
Lys Gly Arg His Asn Ser Glu Lys Pro Pro Glu Pro Val Lys Pro
                 65                  70                  75
Glu Val Lys Thr Thr Glu Lys Lys Glu Leu Cys Glu Leu Lys Pro
                 80                  85                  90
Lys Phe Gln Glu His Ile Ile Gln Ala Pro Lys Pro Val Glu Ala
                 95                 100                 105
Ile Lys Arg Pro Ser Pro Asp Glu Pro Met Thr Asn Leu Glu Leu
```

```
                110                 115                 120
Lys Ile Ser Ala Ser Leu Lys Gln Ala Leu Asp Lys Leu Lys Leu
                125                 130                 135
Ser Ser Gly Asn Glu Glu Asn Lys Lys Glu Glu Asp Asn Asp Glu
                140                 145                 150
Ile Lys Ile Gly Thr Ser Cys Lys Asn Gly Gly Cys Ser Lys Thr
                155                 160                 165
Tyr Gln Gly Leu Glu Ser Leu Glu Glu Val Cys Val Tyr His Ser
                170                 175                 180
Gly Val Pro Ile Phe His Glu Gly Met Lys Tyr Trp Ser Cys Cys
                185                 190                 195
Arg Arg Lys Thr Ser Asp Phe Asn Thr Phe Leu Ala Gln Glu Gly
                200                 205                 210
Cys Thr Lys Gly Lys His Met Trp Thr Lys Lys Asp Ala Gly Lys
                215                 220                 225
Lys Val Val Pro Cys Arg His Asp Trp His Gln Thr Gly Gly Glu
                230                 235                 240
Val Thr Ile Ser Val Tyr Ala Lys Asn Ser Leu Pro Glu Leu Ser
                245                 250                 255
Arg Val Glu Ala Asn Ser Thr Leu Leu Asn Val His Ile Val Phe
                260                 265                 270
Glu Gly Glu Lys Glu Phe Asp Gln Asn Val Lys Leu Trp Gly Val
                275                 280                 285
Ile Asp Val Lys Arg Ser Tyr Val Thr Met Thr Ala Thr Lys Ile
                290                 295                 300
Glu Ile Thr Met Arg Lys Ala Glu Pro Met Gln Trp Ala Ser Leu
                305                 310                 315
Glu Leu Pro Ala Ala Lys Lys Gln Glu Lys Gln Lys Asp Asp Thr
                320                 325                 330
Thr Asp
```

<210> 5
<211> 402
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5519523CD1

<400> 5

```
Met Gln Ser Thr Gly Ser Ser Val Leu Ser Lys Tyr Glu Asp Gln
  1               5                  10                  15
Ile Thr Ile Phe Thr Asp Tyr Leu Glu Glu Tyr Pro Asp Thr Asp
                 20                  25                  30
Glu Leu Val Trp Ile Leu Gly Lys Gln His Leu Leu Lys Thr Glu
                 35                  40                  45
Lys Ser Lys Leu Leu Ser Asp Ile Ser Ala Arg Leu Trp Phe Thr
                 50                  55                  60
Tyr Arg Arg Lys Phe Ser Pro Ile Gly Gly Thr Gly Pro Ser Ser
                 65                  70                  75
Asp Ala Gly Trp Gly Cys Met Leu Arg Cys Gly Gln Met Met Leu
                 80                  85                  90
Ala Gln Ala Leu Ile Cys Arg His Leu Gly Arg Asp Trp Ser Trp
                 95                 100                 105
```

```
Glu Lys Gln Lys Glu Gln Pro Lys Glu Tyr Gln Arg Ile Leu Gln
                110                 115                 120
Cys Phe Leu Asp Arg Lys Asp Cys Cys Tyr Ser Ile His Gln Met
                125                 130                 135
Ala Gln Met Gly Val Gly Glu Gly Lys Ser Ile Gly Glu Trp Phe
                140                 145                 150
Gly Pro Asn Thr Val Ala Gln Val Leu Lys Lys Leu Ala Leu Phe
                155                 160                 165
Asp Glu Trp Asn Ser Leu Ala Val Tyr Val Ser Met Asp Asn Thr
                170                 175                 180
Val Val Ile Glu Asp Ile Lys Lys Met Cys Arg Val Leu Pro Leu
                185                 190                 195
Ser Ala Asp Thr Ala Gly Asp Arg Pro Pro Asp Ser Leu Thr Ala
                200                 205                 210
Ser Asn Gln Ser Lys Gly Thr Ser Ala Tyr Cys Ser Ala Trp Lys
                215                 220                 225
Pro Leu Leu Leu Ile Val Pro Leu Arg Leu Gly Ile Asn Gln Ile
                230                 235                 240
Asn Pro Val Tyr Val Asp Ala Phe Lys Glu Cys Phe Lys Met Pro
                245                 250                 255
Gln Ser Leu Gly Ala Leu Gly Gly Lys Pro Asn Asn Ala Tyr Tyr
                260                 265                 270
Phe Ile Gly Phe Leu Gly Asp Glu Leu Ile Phe Leu Asp Pro His
                275                 280                 285
Thr Thr Gln Thr Phe Val Asp Thr Glu Glu Asn Gly Thr Val Asn
                290                 295                 300
Asp Gln Thr Phe His Cys Leu Gln Ser Pro Gln Arg Met Asn Ile
                305                 310                 315
Leu Asn Leu Asp Pro Ser Val Ala Leu Gly Phe Phe Cys Lys Glu
                320                 325                 330
Glu Lys Asp Phe Asp Asn Trp Cys Ser Leu Val Gln Lys Glu Ile
                335                 340                 345
Leu Lys Glu Asn Leu Arg Met Phe Glu Leu Val Gln Lys His Pro
                350                 355                 360
Ser His Trp Pro Pro Phe Val Pro Pro Ala Lys Pro Glu Val Thr
                365                 370                 375
Thr Thr Gly Ala Glu Phe Ile Asp Ser Thr Glu Gln Leu Glu Glu
                380                 385                 390
Phe Asp Leu Glu Glu Asp Phe Glu Ile Leu Ser Val
                395                 400

<210> 6
<211> 589
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1760208CD1

<400> 6
Met Thr Gly Leu Leu Lys Arg Lys Phe Asp Gln Leu Asp Glu Asp
  1               5                  10                  15
Asn Ser Ser Val Ser Ser Ser Ser Ser Ser Ser Gly Cys Gln Ser
                 20                  25                  30
Arg Ser Cys Ser Pro Ser Ser Ser Val Ser Arg Ala Trp Asp Ser
```

```
                35                  40                  45
Glu Glu Glu Gly Pro Trp Asp Gln Met Pro Leu Pro Asp Arg Asp
                50                  55                  60
Phe Cys Gly Pro Arg Ser Phe Thr Pro Leu Ser Ile Leu Lys Arg
                65                  70                  75
Ala Arg Arg Glu Arg Pro Gly Arg Val Ala Phe Asp Gly Ile Thr
                80                  85                  90
Val Phe Tyr Phe Pro Arg Cys Gln Gly Phe Thr Ser Val Pro Ser
                95                 100                 105
Arg Gly Gly Cys Thr Leu Gly Met Ala Leu Arg His Ser Ala Cys
               110                 115                 120
Arg Arg Phe Ser Leu Ala Glu Phe Ala Gln Glu Gln Ala Arg Ala
               125                 130                 135
Arg His Glu Lys Leu Arg Gln Arg Leu Lys Glu Glu Lys Leu Glu
               140                 145                 150
Met Leu Gln Trp Lys Leu Ser Ala Ala Gly Val Pro Gln Ala Glu
               155                 160                 165
Ala Gly Leu Pro Pro Val Val Asp Ala Ile Asp Asp Ala Ser Val
               170                 175                 180
Glu Glu Asp Leu Ala Val Ala Val Ala Gly Gly Arg Leu Glu Glu
               185                 190                 195
Val Ser Phe Leu Gln Pro Tyr Pro Ala Arg Arg Arg Arg Ala Leu
               200                 205                 210
Leu Arg Ala Ser Gly Val Arg Arg Ile Asp Arg Glu Glu Lys Arg
               215                 220                 225
Glu Leu Gln Ala Leu Arg Gln Ser Arg Glu Asp Cys Gly Cys His
               230                 235                 240
Cys Asp Arg Ile Cys Asp Pro Glu Thr Cys Ser Cys Ser Leu Ala
               245                 250                 255
Gly Ile Lys Cys Gln Met Asp His Thr Ala Phe Pro Cys Gly Cys
               260                 265                 270
Cys Arg Glu Gly Cys Glu Asn Pro Met Gly Arg Val Glu Phe Asn
               275                 280                 285
Gln Ala Arg Val Gln Thr His Phe Ile His Thr Leu Thr Arg Leu
               290                 295                 300
Gln Leu Glu Gln Glu Ala Glu Ser Phe Arg Glu Leu Glu Ala Pro
               305                 310                 315
Ala Gln Gly Ser Pro Pro Ser Pro Gly Glu Glu Ala Leu Val Pro
               320                 325                 330
Thr Phe Pro Leu Ala Lys Pro Pro Met Asn Asn Glu Leu Gly Asp
               335                 340                 345
Asn Ser Cys Ser Ser Asp Met Thr Asp Ser Ser Thr Ala Ser Ser
               350                 355                 360
Ser Ala Ser Gly Thr Ser Glu Ala Pro Asp Cys Pro Thr His Pro
               365                 370                 375
Gly Leu Pro Gly Pro Gly Phe Gln Pro Gly Val Asp Asp Asp Ser
               380                 385                 390
Leu Ala Arg Ile Leu Ser Phe Ser Asp Ser Asp Phe Gly Gly Glu
               395                 400                 405
Glu Glu Glu Glu Glu Glu Gly Ser Val Gly Asn Leu Asp Asn Leu
               410                 415                 420
Ser Cys Phe His Pro Ala Asp Ile Phe Gly Thr Ser Asp Pro Gly
               425                 430                 435
Gly Leu Ala Ser Trp Thr His Ser Tyr Ser Gly Cys Ser Phe Thr
               440                 445                 450
Ser Gly Ile Leu Asp Glu Asn Ala Asn Leu Asp Ala Ser Cys Phe
```

```
                455                 460                 465
Leu Asn Gly Gly Leu Glu Gly Ser Arg Glu Gly Ser Leu Pro Gly
                470                 475                 480
Thr Ser Val Pro Pro Ser Met Asp Ala Gly Arg Ser Ser Ser Val
                485                 490                 495
Asp Leu Ser Leu Ser Ser Cys Asp Ser Phe Glu Leu Leu Gln Ala
                500                 505                 510
Leu Pro Asp Tyr Ser Leu Gly Pro His Tyr Thr Ser Gln Lys Val
                515                 520                 525
Ser Asp Ser Leu Asp Asn Ile Glu Ala Pro His Phe Pro Leu Pro
                530                 535                 540
Gly Leu Ser Pro Pro Gly Asp Ala Ser Ser Cys Phe Leu Glu Ser
                545                 550                 555
Leu Met Gly Phe Ser Glu Pro Ala Ala Glu Ala Leu Asp Pro Phe
                560                 565                 570
Ile Asp Ser Gln Phe Glu Asp Thr Val Pro Ala Ser Leu Met Glu
                575                 580                 585
Pro Val Pro Val
```

<210> 7
<211> 741
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1900132CD1

<400> 7

```
Met Ala Lys Leu Asn Tyr Val Glu Gly Asp Tyr Lys Glu Ala Leu
  1                 5                  10                  15
Asn Ile Tyr Ala Arg Val Gly Leu Asp Asp Leu Pro Leu Thr Ala
                 20                  25                  30
Val Pro Pro Tyr Arg Leu Arg Val Ile Ala Glu Ala Tyr Ala Thr
                 35                  40                  45
Lys Gly Leu Cys Leu Glu Lys Leu Pro Ile Ser Ser Ser Thr Ser
                 50                  55                  60
Asn Leu His Val Asp Arg Glu Gln Asp Val Ile Thr Cys Tyr Glu
                 65                  70                  75
Lys Ala Gly Asp Ile Ala Leu Leu Tyr Leu Gln Glu Ile Glu Arg
                 80                  85                  90
Val Ile Leu Ser Asn Ile Gln Asn Arg Ser Pro Lys Pro Gly Pro
                 95                 100                 105
Ala Pro His Asp Gln Glu Leu Gly Phe Phe Leu Glu Thr Gly Leu
                110                 115                 120
Gln Arg Ala His Val Leu Tyr Phe Lys Asn Gly Asn Leu Thr Arg
                125                 130                 135
Gly Val Gly Arg Phe Arg Glu Leu Leu Arg Ala Val Glu Thr Arg
                140                 145                 150
Thr Thr Gln Asn Leu Arg Met Thr Ile Ala Arg Gln Leu Ala Glu
                155                 160                 165
Ile Leu Leu Arg Gly Met Cys Glu Gln Ser Tyr Trp Asn Pro Leu
                170                 175                 180
Glu Asp Pro Pro Cys Gln Ser Pro Leu Asp Asp Pro Leu Arg Lys
                185                 190                 195
```

```
Gly Ala Asn Thr Lys Thr Tyr Thr Leu Thr Arg Arg Ala Arg Val
                200                 205                 210
Tyr Ser Gly Glu Asn Ile Phe Cys Pro Gln Glu Asn Thr Glu Glu
                215                 220                 225
Ala Leu Leu Leu Leu Leu Ile Ser Glu Ser Met Ala Asn Arg Asp
                230                 235                 240
Ala Val Leu Ser Arg Ile Pro Glu His Lys Ser Asp Arg Leu Ile
                245                 250                 255
Ser Leu Gln Ser Ala Ser Val Val Tyr Asp Leu Leu Thr Ile Ala
                260                 265                 270
Leu Gly Arg Arg Gly Gln Tyr Glu Met Leu Ser Glu Cys Leu Glu
                275                 280                 285
Arg Ala Met Lys Phe Ala Phe Glu Glu Phe His Leu Trp Tyr Gln
                290                 295                 300
Phe Ala Leu Ser Leu Met Ala Ala Gly Lys Ser Ala Arg Ala Val
                305                 310                 315
Lys Val Leu Lys Glu Cys Ile Arg Leu Lys Pro Asp Asp Ala Thr
                320                 325                 330
Ile Pro Leu Leu Ala Ala Lys Leu Cys Met Gly Ser Leu His Trp
                335                 340                 345
Leu Glu Glu Ala Glu Lys Phe Ala Lys Thr Val Val Asp Val Gly
                350                 355                 360
Glu Lys Thr Ser Glu Phe Lys Ala Lys Gly Tyr Leu Ala Leu Gly
                365                 370                 375
Leu Thr Tyr Ser Leu Gln Ala Thr Asp Ala Ser Leu Arg Gly Met
                380                 385                 390
Gln Glu Val Leu Gln Arg Lys Ala Leu Leu Ala Phe Gln Arg Ala
                395                 400                 405
His Ser Leu Ser Pro Thr Asp His Gln Ala Ala Phe Tyr Leu Ala
                410                 415                 420
Leu Gln Leu Ala Ile Ser Arg Gln Ile Pro Glu Ala Leu Gly Tyr
                425                 430                 435
Val Arg Gln Ala Leu Gln Leu Gln Gly Asp Asp Ala Asn Ser Leu
                440                 445                 450
His Leu Leu Ala Leu Leu Leu Ser Ala Gln Lys His Tyr His Asp
                455                 460                 465
Ala Leu Asn Ile Ile Asp Met Ala Leu Ser Glu Tyr Pro Glu Asn
                470                 475                 480
Phe Ile Leu Leu Phe Ser Lys Val Lys Leu Gln Ser Leu Cys Arg
                485                 490                 495
Gly Pro Asp Glu Ala Leu Leu Thr Cys Lys His Met Leu Gln Ile
                500                 505                 510
Trp Lys Ser Cys Tyr Asn Leu Thr Asn Pro Ser Asp Ser Gly Arg
                515                 520                 525
Gly Ser Ser Leu Leu Asp Arg Thr Ile Ala Asp Arg Arg Gln Leu
                530                 535                 540
Asn Thr Ile Thr Leu Pro Asp Phe Ser Asp Pro Glu Thr Gly Ser
                545                 550                 555
Val His Ala Thr Ser Val Ala Ala Ser Arg Val Glu Gln Ala Leu
                560                 565                 570
Ser Glu Val Ala Ser Ser Leu Gln Ser Ser Ala Pro Lys Gln Gly
                575                 580                 585
Pro Leu His Pro Trp Met Thr Leu Ala Gln Ile Trp Leu His Ala
                590                 595                 600
Ala Glu Val Tyr Ile Gly Ile Gly Lys Pro Ala Glu Ala Thr Ala
                605                 610                 615
```

```
Cys Thr Gln Glu Ala Ala Asn Leu Phe Pro Met Ser His Asn Val
                620                 625                 630
Leu Tyr Met Arg Gly Gln Ile Ala Glu Leu Arg Gly Ser Met Asp
                635                 640                 645
Glu Ala Arg Arg Trp Tyr Glu Glu Ala Leu Ala Ile Ser Pro Thr
                650                 655                 660
His Val Lys Ser Met Gln Arg Leu Ala Leu Ile Leu His Gln Leu
                665                 670                 675
Gly Arg Tyr Ser Leu Ala Glu Lys Ile Leu Arg Asp Ala Val Gln
                680                 685                 690
Val Asn Ser Thr Ala His Glu Val Trp Asn Gly Leu Gly Glu Val
                695                 700                 705
Leu Gln Ala Gln Gly Asn Asp Ala Ala Ala Thr Glu Cys Phe Leu
                710                 715                 720
Thr Ala Leu Glu Leu Glu Ala Ser Ser Pro Ala Val Pro Phe Thr
                725                 730                 735
Ile Ile Pro Arg Val Leu
                740

<210> 8
<211> 227
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7487551CD1

<400> 8
Met Gly Thr Pro Thr Pro Asp Thr His Pro Ile Leu Arg Thr Glu
  1               5                  10                  15
Trp Gly Trp Glu Glu Pro Val Ala Lys Gly Gly Glu Glu Gly Arg
                20                  25                  30
Ala Glu Ser Arg Trp Gly Pro Pro Leu Val Ala Ser Ser Leu His
                35                  40                  45
Gly Pro Arg Leu Gln Pro Thr Trp Val Leu Gly Val Gly Gly Ser
                50                  55                  60
Ser Thr Trp Ala Met Ala Glu Asp Arg Pro Gln Gln Pro Gln Leu
                65                  70                  75
Asp Met Pro Leu Val Leu Asp Gln Gly Leu Thr Arg Gln Met Arg
                80                  85                  90
Leu Arg Val Glu Ser Leu Lys Gln Arg Gly Glu Lys Arg Gln Asp
                95                 100                 105
Gly Glu Lys Leu Leu Gln Pro Ala Glu Ser Val Tyr Arg Leu Asn
                110                 115                 120
Phe Thr Gln Gln Gln Arg Leu Gln Phe Glu Arg Trp Asn Val Val
                125                 130                 135
Leu Asp Lys Pro Gly Lys Val Thr Ile Thr Gly Thr Ser Gln Asn
                140                 145                 150
Trp Thr Pro Asp Leu Thr Asn Leu Met Thr Arg Gln Leu Leu Asp
                155                 160                 165
Pro Thr Ala Ile Phe Trp Arg Lys Glu Asp Ser Asp Ala Ile Asp
                170                 175                 180
Trp Asn Glu Ala Asp Ala Leu Glu Phe Gly Glu Arg Leu Ser Asp
                185                 190                 195
Leu Ala Lys Ile Arg Lys Val Met Tyr Phe Leu Val Thr Phe Gly
```

```
                200                 205                 210
Glu Gly Val Glu Pro Ala Asn Leu Lys Ala Ser Val Val Phe Asn
                215                 220                 225
Gln Leu
```

<210> 9
<211> 261
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1871014CD1

<400> 9

```
Met Ala Ala Ala Val Ala Gly Met Leu Arg Gly Gly Leu Leu Pro
  1               5                  10                  15
Gln Ala Gly Arg Leu Pro Thr Leu Gln Thr Val Arg Tyr Gly Ser
                20                  25                  30
Lys Ala Val Thr Arg His Arg Arg Val Met His Phe Gln Arg Gln
                35                  40                  45
Lys Leu Met Ala Val Thr Glu Tyr Ile Pro Pro Lys Pro Ala Ile
                50                  55                  60
His Pro Ser Cys Leu Pro Ser Pro Pro Ser Pro Pro Gln Glu Glu
                65                  70                  75
Ile Gly Leu Ile Arg Leu Leu Arg Arg Glu Ile Ala Ala Val Phe
                80                  85                  90
Gln Asp Asn Arg Met Ile Ala Val Cys Gln Asn Val Ala Leu Ser
                95                 100                 105
Ala Glu Asp Lys Leu Leu Met Arg His Gln Leu Arg Lys His Lys
                110                115                 120
Ile Leu Met Lys Val Phe Pro Asn Gln Val Leu Lys Pro Phe Leu
                125                130                 135
Glu Asp Ser Lys Tyr Gln Asn Leu Leu Pro Leu Phe Val Gly His
                140                145                 150
Asn Met Leu Leu Val Ser Glu Glu Pro Lys Val Lys Glu Met Val
                155                160                 165
Arg Ile Leu Arg Thr Val Pro Phe Leu Pro Leu Leu Gly Gly Cys
                170                175                 180
Ile Asp Asp Thr Ile Leu Ser Arg Gln Gly Phe Ile Asn Tyr Ser
                185                190                 195
Lys Leu Pro Ser Leu Pro Leu Val Gln Gly Glu Leu Val Gly Gly
                200                205                 210
Leu Thr Cys Leu Thr Ala Gln Thr His Ser Leu Leu Gln His Gln
                215                220                 225
Pro Leu Gln Leu Thr Thr Leu Leu Asp Gln Tyr Ile Arg Glu Gln
                230                235                 240
Arg Glu Lys Asp Ser Val Met Ser Ala Asn Gly Lys Pro Asp Pro
                245                250                 255
Asp Thr Val Pro Asp Ser
                260
```

<210> 10
<211> 1461
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2903166CD1

<400> 10
Met Glu Ala Arg Ser Arg Ser Ala Glu Glu Leu Arg Arg Ala Glu
1 5 10 15
Leu Val Glu Ile Ile Val Glu Thr Glu Ala Gln Thr Gly Val Ser
20 25 30
Gly Ile Asn Val Ala Gly Gly Gly Lys Glu Gly Ile Phe Val Arg
35 40 45
Glu Leu Arg Glu Asp Ser Pro Ala Ala Arg Ser Leu Ser Leu Gln
50 55 60
Glu Gly Asp Gln Leu Leu Ser Ala Arg Val Phe Phe Glu Asn Phe
65 70 75
Lys Tyr Glu Asp Ala Leu Arg Leu Leu Gln Cys Ala Glu Pro Tyr
80 85 90
Lys Val Ser Phe Cys Leu Lys Arg Thr Val Pro Thr Gly Asp Leu
95 100 105
Ala Leu Arg Pro Gly Thr Val Ser Gly Tyr Glu Ile Lys Gly Pro
110 115 120
Arg Ala Lys Val Ala Lys Leu Asn Ile Gln Ser Leu Ser Pro Val
125 130 135
Lys Lys Lys Lys Met Val Pro Gly Ala Leu Gly Val Pro Ala Asp
140 145 150
Leu Ala Pro Val Asp Val Glu Phe Ser Phe Pro Lys Phe Ser Arg
155 160 165
Leu Arg Arg Gly Leu Lys Ala Glu Ala Val Lys Gly Pro Val Pro
170 175 180
Ala Ala Pro Ala Arg Arg Arg Leu Gln Leu Pro Arg Leu Arg Val
185 190 195
Arg Glu Val Ala Glu Glu Ala Gln Ala Ala Arg Leu Ala Ala Ala
200 205 210
Ala Pro Pro Pro Arg Lys Ala Lys Val Glu Ala Glu Val Ala Ala
215 220 225
Gly Ala Arg Phe Thr Ala Pro Gln Val Glu Leu Val Gly Pro Arg
230 235 240
Leu Pro Gly Ala Glu Val Gly Val Pro Gln Val Ser Ala Pro Lys
245 250 255
Ala Ala Pro Ser Ala Glu Ala Ala Gly Gly Phe Ala Leu His Leu
260 265 270
Pro Thr Leu Gly Leu Gly Ala Pro Ala Pro Pro Ala Val Glu Ala
275 280 285
Pro Ala Val Gly Ile Gln Val Pro Gln Val Glu Leu Pro Ala Leu
290 295 300
Pro Ser Leu Pro Thr Leu Pro Thr Leu Pro Cys Leu Glu Thr Arg
305 310 315
Glu Gly Ala Val Ser Val Val Val Pro Thr Leu Asp Val Ala Ala
320 325 330
Pro Thr Val Gly Val Asp Leu Ala Leu Pro Gly Ala Glu Val Glu
335 340 345
Ala Arg Gly Glu Ala Pro Glu Val Ala Leu Lys Met Pro Arg Leu
350 355 360
Ser Phe Pro Arg Phe Gly Ala Arg Ala Lys Glu Val Ala Glu Ala

```
                365                 370                 375
Lys Val Ala Lys Val Ser Pro Glu Ala Arg Val Lys Gly Pro Arg
                380                 385                 390
Leu Arg Met Pro Thr Phe Gly Leu Ser Leu Leu Glu Pro Arg Pro
                395                 400                 405
Ala Ala Pro Glu Val Val Glu Ser Lys Leu Lys Leu Pro Thr Ile
                410                 415                 420
Lys Met Pro Ser Leu Gly Ile Gly Val Ser Gly Pro Glu Val Lys
                425                 430                 435
Val Pro Lys Gly Pro Glu Val Lys Leu Pro Lys Ala Pro Glu Val
                440                 445                 450
Lys Leu Pro Lys Val Pro Glu Ala Ala Leu Pro Glu Val Arg Leu
                455                 460                 465
Pro Glu Val Glu Leu Pro Lys Val Ser Glu Met Lys Leu Pro Lys
                470                 475                 480
Val Pro Glu Met Ala Val Pro Glu Val Arg Leu Pro Glu Val Glu
                485                 490                 495
Leu Pro Lys Val Ser Glu Met Lys Leu Pro Lys Val Pro Glu Met
                500                 505                 510
Ala Val Pro Glu Val Arg Leu Pro Glu Val Gln Leu Leu Lys Val
                515                 520                 525
Ser Glu Met Lys Leu Pro Lys Val Pro Glu Met Ala Val Pro Glu
                530                 535                 540
Val Arg Leu Pro Glu Val Gln Leu Pro Lys Val Ser Glu Met Lys
                545                 550                 555
Leu Pro Glu Val Ser Glu Val Ala Val Pro Glu Val Arg Leu Pro
                560                 565                 570
Glu Val Gln Leu Pro Lys Val Pro Glu Met Lys Val Pro Glu Met
                575                 580                 585
Lys Leu Pro Lys Val Pro Glu Met Lys Leu Pro Glu Met Lys Leu
                590                 595                 600
Pro Glu Val Gln Leu Pro Lys Val Pro Glu Met Ala Val Pro Asp
                605                 610                 615
Val His Leu Pro Glu Val Gln Leu Pro Lys Val Pro Glu Met Lys
                620                 625                 630
Leu Pro Glu Met Lys Leu Pro Glu Val Lys Leu Pro Lys Val Pro
                635                 640                 645
Glu Met Ala Val Pro Asp Val His Leu Pro Glu Val Gln Leu Pro
                650                 655                 660
Lys Val Pro Glu Met Lys Leu Pro Lys Met Pro Glu Met Ala Val
                665                 670                 675
Pro Glu Val Arg Leu Pro Glu Val Gln Leu Pro Lys Val Ser Glu
                680                 685                 690
Met Lys Leu Pro Lys Val Pro Glu Met Ala Val Pro Asp Val His
                695                 700                 705
Leu Pro Glu Val Gln Leu Pro Lys Val Cys Glu Met Lys Val Pro
                710                 715                 720
Asp Met Lys Leu Pro Glu Ile Lys Leu Pro Lys Val Pro Glu Met
                725                 730                 735
Ala Val Pro Asp Val His Leu Pro Glu Val Gln Leu Pro Lys Val
                740                 745                 750
Ser Glu Ile Arg Leu Pro Glu Met Gln Val Pro Lys Val Pro Asp
                755                 760                 765
Val His Leu Pro Lys Ala Pro Glu Val Lys Leu Pro Arg Ala Pro
                770                 775                 780
Glu Val Gln Leu Lys Ala Thr Lys Ala Glu Gln Ala Glu Gly Met
```

```
                785                 790                 795
Glu Phe Gly Phe Lys Met Pro Lys Met Thr Met Pro Lys Leu Gly
                800                 805                 810
Arg Ala Glu Ser Pro Ser Arg Gly Lys Pro Gly Glu Ala Gly Ala
                815                 820                 825
Glu Val Ser Gly Lys Leu Val Thr Leu Pro Cys Leu Gln Pro Glu
                830                 835                 840
Val Asp Gly Glu Ala His Val Gly Val Pro Ser Leu Thr Leu Pro
                845                 850                 855
Ser Val Glu Leu Asp Leu Pro Gly Ala Leu Gly Leu Gln Gly Gln
                860                 865                 870
Val Pro Ala Ala Lys Met Gly Lys Gly Glu Arg Ala Glu Gly Pro
                875                 880                 885
Glu Val Ala Ala Gly Val Arg Glu Val Gly Phe Arg Val Pro Ser
                890                 895                 900
Val Glu Ile Val Thr Pro Gln Leu Pro Ala Val Glu Ile Glu Glu
                905                 910                 915
Gly Arg Leu Glu Met Ile Glu Thr Lys Val Lys Pro Ser Ser Lys
                920                 925                 930
Phe Ser Leu Pro Lys Phe Gly Leu Ser Gly Pro Lys Val Ala Lys
                935                 940                 945
Ala Glu Ala Glu Gly Ala Gly Arg Ala Thr Lys Leu Lys Val Ser
                950                 955                 960
Lys Phe Ala Ile Ser Leu Pro Lys Ala Arg Val Gly Ala Glu Ala
                965                 970                 975
Glu Ala Lys Gly Ala Gly Glu Ala Gly Leu Leu Pro Ala Leu Asp
                980                 985                 990
Leu Ser Ile Pro Gln Leu Ser Leu Asp Ala His Leu Pro Ser Gly
                995                1000                1005
Lys Val Glu Val Ala Gly Ala Asp Leu Lys Phe Lys Gly Pro Arg
               1010                1015                1020
Phe Ala Leu Pro Lys Phe Gly Val Arg Gly Arg Asp Thr Glu Ala
               1025                1030                1035
Ala Glu Leu Val Pro Gly Val Ala Glu Leu Glu Gly Lys Gly Trp
               1040                1045                1050
Gly Trp Asp Gly Arg Val Lys Met Pro Lys Leu Lys Met Pro Ser
               1055                1060                1065
Phe Gly Leu Ala Arg Gly Lys Glu Ala Glu Val Gln Gly Asp Arg
               1070                1075                1080
Ala Ser Pro Gly Glu Lys Ala Glu Ser Thr Ala Val Gln Leu Lys
               1085                1090                1095
Ile Pro Glu Val Glu Leu Val Thr Leu Gly Ala Gln Glu Glu Gly
               1100                1105                1110
Arg Ala Glu Gly Ala Val Ala Val Ser Gly Met Gln Leu Ser Gly
               1115                1120                1125
Leu Lys Val Ser Thr Ala Arg Gln Val Val Thr Glu Gly His Asp
               1130                1135                1140
Ala Gly Leu Arg Met Pro Pro Leu Gly Ile Ser Leu Pro Gln Val
               1145                1150                1155
Glu Leu Thr Gly Phe Gly Glu Ala Gly Thr Pro Gly Gln Gln Ala
               1160                1165                1170
Gln Ser Thr Val Pro Ser Ala Glu Gly Thr Ala Gly Tyr Arg Val
               1175                1180                1185
Gln Val Pro Gln Val Thr Leu Ser Leu Pro Gly Ala Gln Val Ala
               1190                1195                1200
Gly Gly Glu Leu Leu Val Gly Glu Gly Val Phe Lys Met Pro Thr
```

```
                1205                1210                1215
Val Thr Val Pro Gln Leu Glu Leu Asp Val Gly Leu Ser Arg Glu
                1220                1225                1230
Ala Gln Ala Gly Glu Ala Ala Thr Gly Glu Gly Gly Leu Arg Leu
                1235                1240                1245
Lys Leu Pro Thr Leu Gly Ala Arg Ala Arg Val Gly Gly Glu Gly
                1250                1255                1260
Ala Glu Glu Gln Pro Pro Gly Ala Glu Arg Thr Phe Cys Leu Ser
                1265                1270                1275
Leu Pro Asp Val Glu Leu Ser Pro Ser Gly Gly Asn His Ala Glu
                1280                1285                1290
Tyr Gln Val Ala Glu Gly Glu Gly Glu Ala Gly His Lys Leu Lys
                1295                1300                1305
Val Arg Leu Pro Arg Phe Gly Leu Val Arg Ala Lys Glu Gly Ala
                1310                1315                1320
Glu Glu Gly Glu Lys Ala Lys Ser Pro Lys Leu Arg Leu Pro Arg
                1325                1330                1335
Val Gly Phe Ser Gln Ser Glu Met Val Thr Gly Glu Gly Ser Pro
                1340                1345                1350
Ser Pro Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Gly Ser Gly Glu
                1355                1360                1365
Gly Ala Ser Gly Arg Arg Gly Arg Val Arg Val Arg Leu Pro Arg
                1370                1375                1380
Val Gly Leu Ala Ala Pro Ser Lys Ala Ser Arg Gly Gln Glu Gly
                1385                1390                1395
Asp Ala Ala Pro Lys Ser Pro Val Arg Glu Lys Ser Pro Lys Phe
                1400                1405                1410
Arg Phe Pro Arg Val Ser Leu Ser Pro Lys Ala Arg Ser Gly Ser
                1415                1420                1425
Gly Asp Gln Glu Glu Gly Gly Leu Arg Val Arg Leu Pro Ser Val
                1430                1435                1440
Gly Phe Ser Glu Thr Gly Ala Pro Gly Pro Ala Arg Met Glu Gly
                1445                1450                1455
Ala Gln Ala Ala Ala Val
                1460

<210> 11
<211> 657
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1723804CD1

<400> 11
Met Glu Met Glu Thr Thr Glu Pro Glu Pro Asp Cys Val Val Gln
  1               5                  10                  15
Pro Pro Ser Pro Pro Asp Asp Phe Ser Cys Gln Met Arg Leu Ser
                20                  25                  30
Glu Lys Ile Thr Pro Leu Lys Thr Cys Phe Lys Lys Lys Asp Gln
                35                  40                  45
Lys Arg Leu Gly Thr Gly Thr Leu Arg Ser Leu Arg Pro Ile Leu
                50                  55                  60
Asn Thr Leu Leu Glu Ser Gly Ser Leu Asp Gly Val Phe Arg Ser
                65                  70                  75
```

```
Arg Asn Gln Ser Thr Asp Glu Asn Ser Leu His Glu Pro Met Met
                 80                  85                  90
Lys Lys Ala Met Glu Ile Asn Ser Ser Cys Pro Pro Ala Glu Asn
                 95                 100                 105
Asn Met Ser Val Leu Ile Pro Asp Arg Thr Asn Val Gly Asp Gln
                110                 115                 120
Ile Pro Glu Ala His Pro Ser Thr Glu Ala Pro Glu Arg Val Val
                125                 130                 135
Pro Ile Gln Asp His Ser Phe Pro Ser Glu Thr Leu Ser Gly Thr
                140                 145                 150
Val Ala Asp Ser Thr Pro Ala His Phe Gln Thr Asp Leu Leu His
                155                 160                 165
Pro Val Ser Ser Asp Val Pro Thr Ser Pro Asp Cys Leu Asp Lys
                170                 175                 180
Val Ile Asp Tyr Val Pro Gly Ile Phe Gln Glu Asn Ser Phe Thr
                185                 190                 195
Ile Gln Tyr Ile Leu Asp Thr Ser Asp Lys Leu Ser Thr Glu Leu
                200                 205                 210
Phe Gln Asp Lys Ser Glu Glu Ala Ser Leu Asp Leu Val Phe Glu
                215                 220                 225
Leu Val Asn Gln Leu Gln Tyr His Thr His Gln Glu Asn Gly Ile
                230                 235                 240
Glu Ile Cys Met Asp Phe Leu Gln Gly Thr Cys Ile Tyr Gly Arg
                245                 250                 255
Asp Cys Leu Lys His His Thr Val Leu Pro Tyr His Trp Gln Ile
                260                 265                 270
Lys Arg Thr Thr Thr Gln Lys Trp Gln Ser Val Phe Asn Asp Ser
                275                 280                 285
Gln Glu His Leu Glu Arg Phe Tyr Cys Asn Pro Glu Asn Asp Arg
                290                 295                 300
Met Arg Met Lys Tyr Gly Gly Gln Glu Phe Trp Ala Asp Leu Asn
                305                 310                 315
Ala Met Asn Val Tyr Glu Thr Thr Glu Phe Asp Gln Leu Arg Arg
                320                 325                 330
Leu Ser Thr Pro Pro Ser Ser Asn Val Asn Ser Ile Tyr His Thr
                335                 340                 345
Val Trp Lys Phe Phe Cys Arg Asp His Phe Gly Trp Arg Glu Tyr
                350                 355                 360
Pro Glu Ser Val Ile Arg Leu Ile Glu Glu Ala Asn Ser Arg Gly
                365                 370                 375
Leu Lys Glu Val Arg Phe Met Met Trp Asn Asn His Tyr Ile Leu
                380                 385                 390
His Asn Ser Phe Phe Arg Arg Glu Ile Lys Arg Arg Pro Leu Phe
                395                 400                 405
Arg Ser Cys Phe Ile Leu Leu Pro Tyr Leu Gln Thr Leu Gly Gly
                410                 415                 420
Val Pro Thr Gln Ala Pro Pro Pro Leu Glu Ala Thr Ser Ser Ser
                425                 430                 435
Gln Ile Ile Cys Pro Asp Gly Val Thr Ser Ala Asn Phe Tyr Pro
                440                 445                 450
Glu Thr Trp Val Tyr Met His Pro Ser Gln Asp Phe Ile Gln Val
                455                 460                 465
Pro Val Ser Ala Glu Asp Lys Ser Tyr Arg Ile Ile Tyr Asn Leu
                470                 475                 480
Phe His Lys Thr Val Pro Glu Phe Lys Tyr Arg Ile Leu Gln Ile
                485                 490                 495
```

```
Leu Arg Val Gln Asn Gln Phe Leu Trp Glu Lys Tyr Lys Arg Lys
                500                 505                 510
Lys Glu Tyr Met Asn Arg Lys Met Phe Gly Arg Asp Arg Ile Ile
                515                 520                 525
Asn Glu Arg His Leu Phe His Gly Thr Ser Gln Asp Val Val Asp
                530                 535                 540
Gly Ile Cys Lys His Asn Phe Asp Pro Arg Val Cys Gly Lys His
                545                 550                 555
Ala Thr Met Phe Gly Gln Gly Ser Tyr Phe Ala Lys Lys Ala Ser
                560                 565                 570
Tyr Ser His Asn Phe Ser Lys Lys Ser Ser Lys Gly Val His Phe
                575                 580                 585
Met Phe Leu Ala Lys Val Leu Thr Gly Arg Tyr Thr Met Gly Ser
                590                 595                 600
His Gly Met Arg Arg Pro Pro Pro Val Asn Pro Gly Ser Val Thr
                605                 610                 615
Ser Asp Leu Tyr Asp Ser Cys Val Asp Asn Phe Phe Glu Pro Gln
                620                 625                 630
Ile Phe Val Ile Phe Asn Asp Asp Gln Ser Tyr Pro Tyr Phe Val
                635                 640                 645
Ile Gln Tyr Glu Glu Val Ser Asn Thr Val Ser Ile
                650                 655

<210> 12
<211> 587
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7736769CD1

<400> 12
Met Ala Ala Ala Val Ala Val Ala Ala Ala Ser Arg Arg Gln Ser
  1                 5                  10                  15
Cys Tyr Leu Cys Asp Leu Pro Arg Met Pro Trp Ala Met Ile Trp
                 20                  25                  30
Asp Phe Thr Glu Pro Val Cys Arg Gly Cys Val Asn Tyr Glu Gly
                 35                  40                  45
Ala Asp Arg Val Glu Phe Val Ile Glu Thr Ala Arg Gln Leu Lys
                 50                  55                  60
Arg Ala His Gly Cys Phe Pro Glu Gly Arg Ser Pro Pro Gly Ala
                 65                  70                  75
Ala Ala Ser Ala Ala Ala Lys Pro Pro Pro Leu Ser Ala Lys Asp
                 80                  85                  90
Ile Leu Leu Gln Gln Gln Gln Leu Gly His Gly Gly Pro Glu
                 95                 100                 105
Ala Ala Pro Arg Ala Pro Gln Ala Leu Glu Arg Tyr Pro Leu Ala
                110                 115                 120
Ala Ala Ala Glu Arg Pro Pro Arg Leu Gly Ser Asp Phe Gly Ser
                125                 130                 135
Ser Arg Pro Ala Ala Ser Leu Ala Gln Pro Pro Thr Pro Gln Pro
                140                 145                 150
Pro Pro Val Asn Gly Ile Leu Val Pro Asn Gly Phe Ser Lys Leu
                155                 160                 165
Glu Glu Pro Pro Glu Leu Asn Arg Gln Ser Pro Asn Pro Arg Arg
```

```
                170                 175                 180
Gly His Ala Val Pro Pro Thr Leu Val Pro Leu Met Asn Gly Ser
                185                 190                 195
Ala Thr Pro Leu Pro Thr Ala Leu Gly Leu Gly Gly Arg Ala Ala
                200                 205                 210
Ala Ser Leu Ala Ala Val Ser Gly Thr Ala Ala Ala Ser Leu Gly
                215                 220                 225
Ser Ala Gln Pro Thr Asp Leu Gly Ala His Lys Arg Pro Ala Ser
                230                 235                 240
Val Ser Ser Ser Ala Ala Val Glu His Glu Gln Arg Glu Ala Ala
                245                 250                 255
Ala Lys Glu Lys Gln Pro Pro Pro Pro Ala His Arg Gly Pro Ala
                260                 265                 270
Asp Ser Leu Ser Thr Ala Ala Gly Ala Ala Glu Leu Ser Ala Glu
                275                 280                 285
Gly Ala Gly Lys Ser Arg Gly Ser Gly Glu Gln Asp Trp Val Asn
                290                 295                 300
Arg Pro Lys Thr Val Arg Asp Thr Leu Leu Ala Leu His Gln His
                305                 310                 315
Gly His Ser Gly Pro Phe Glu Ser Lys Phe Lys Lys Glu Pro Ala
                320                 325                 330
Leu Thr Ala Gly Arg Leu Leu Gly Phe Glu Ala Asn Gly Ala Asn
                335                 340                 345
Gly Ser Lys Ala Val Ala Arg Thr Ala Arg Lys Arg Lys Pro Ser
                350                 355                 360
Pro Glu Pro Glu Gly Glu Val Gly Pro Pro Lys Ile Asn Gly Glu
                365                 370                 375
Ala Gln Pro Trp Leu Ser Thr Ser Thr Glu Gly Leu Lys Ile Pro
                380                 385                 390
Met Thr Pro Thr Ser Ser Phe Val Ser Pro Pro Pro Pro Thr Ala
                395                 400                 405
Ser Pro His Ser Asn Arg Thr Thr Pro Pro Glu Ala Ala Gln Asn
                410                 415                 420
Gly Gln Ser Pro Met Ala Ala Leu Ile Leu Val Ala Asp Asn Ala
                425                 430                 435
Gly Gly Ser His Ala Ser Lys Asp Ala Asn Gln Val His Ser Thr
                440                 445                 450
Thr Arg Arg Asn Ser Asn Ser Pro Pro Ser Pro Ser Ser Met Asn
                455                 460                 465
Gln Arg Arg Leu Gly Pro Arg Glu Val Gly Gly Gln Gly Ala Gly
                470                 475                 480
Asn Thr Gly Gly Leu Glu Pro Val His Pro Ala Ser Leu Pro Asp
                485                 490                 495
Ser Ser Leu Ala Thr Ser Ala Pro Leu Cys Cys Thr Leu Cys His
                500                 505                 510
Glu Arg Leu Glu Asp Thr His Phe Val Gln Cys Pro Ser Val Pro
                515                 520                 525
Ser His Lys Phe Cys Phe Pro Cys Ser Arg Gln Ser Ile Lys Gln
                530                 535                 540
Gln Gly Ala Ser Gly Glu Val Tyr Cys Pro Ser Gly Glu Lys Cys
                545                 550                 555
Pro Leu Val Gly Ser Asn Val Pro Trp Ala Phe Met Gln Gly Glu
                560                 565                 570
Ile Ala Thr Ile Leu Ala Gly Asp Val Lys Val Lys Lys Glu Arg
                575                 580                 585
Asp Ser
```

<210> 13
<211> 583
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7492451CD1

<400> 13
Met Ala Ala Ala Ala Val Ser Glu Ser Trp Pro Glu Leu Glu Leu
1 5 10 15
Ala Glu Arg Glu Arg Arg Arg Glu Leu Leu Leu Thr Gly Pro Gly
20 25 30
Leu Glu Glu Arg Val Arg Ala Ala Gly Gly Gln Leu Pro Pro Arg
35 40 45
Leu Phe Thr Leu Pro Leu Leu His Tyr Leu Glu Val Ser Gly Cys
50 55 60
Gly Ser Leu Arg Ala Pro Gly Pro Gly Leu Ala Gln Gly Leu Pro
65 70 75
Gln Leu His Ser Leu Val Leu Arg Arg Asn Ala Leu Gly Pro Gly
80 85 90
Leu Ser Pro Glu Leu Gly Pro Leu Pro Ala Leu Arg Val Leu Asp
95 100 105
Leu Ser Gly Asn Ala Leu Glu Ala Leu Pro Pro Gly Gln Gly Leu
110 115 120
Gly Pro Ala Glu Pro Pro Gly Leu Pro Gln Leu Gln Ser Leu Asn
125 130 135
Leu Ser Gly Asn Arg Leu Arg Glu Leu Pro Ala Asp Leu Ala Arg
140 145 150
Cys Ala Pro Arg Leu Gln Ser Leu Asn Leu Thr Gly Asn Cys Leu
155 160 165
Asp Ser Phe Pro Ala Glu Leu Phe Arg Pro Gly Ala Leu Pro Leu
170 175 180
Leu Ser Glu Leu Ala Ala Ala Asp Asn Cys Leu Arg Glu Leu Ser
185 190 195
Pro Asp Ile Ala His Leu Ala Ser Leu Lys Thr Leu Asp Leu Ser
200 205 210
Asn Asn Gln Leu Ser Glu Ile Pro Ala Glu Leu Ala Asp Cys Pro
215 220 225
Lys Leu Lys Glu Ile Asn Phe Arg Gly Asn Lys Leu Arg Asp Lys
230 235 240
Arg Leu Glu Lys Met Val Ser Gly Cys Gln Thr Arg Ser Ile Leu
245 250 255
Glu Tyr Leu Arg Val Gly Gly Arg Gly Gly Gly Lys Gly Lys Gly
260 265 270
Arg Ala Glu Gly Ser Glu Lys Glu Glu Ser Arg Arg Lys Arg Arg
275 280 285
Glu Arg Lys Gln Arg Arg Glu Gly Gly Asp Gly Glu Glu Gln Asp
290 295 300
Val Gly Asp Ala Gly Arg Leu Leu Leu Arg Val Leu His Val Ser
305 310 315
Glu Asn Pro Val Pro Leu Thr Val Arg Val Ser Pro Glu Val Arg
320 325 330

```
Asp Val Arg Pro Tyr Ile Val Gly Ala Val Val Arg Gly Met Asp
                335                 340                 345
Leu Gln Pro Gly Asn Ala Leu Lys Arg Phe Leu Thr Ser Gln Thr
                350                 355                 360
Lys Leu His Glu Asp Leu Cys Glu Lys Arg Thr Ala Ala Thr Leu
                365                 370                 375
Ala Thr His Glu Leu Arg Ala Val Lys Gly Pro Leu Leu Tyr Cys
                380                 385                 390
Ala Arg Pro Pro Gln Asp Leu Lys Ile Val Pro Leu Gly Arg Lys
                395                 400                 405
Glu Asp Lys Ala Lys Glu Leu Val Arg Gln Leu Gln Leu Glu Ala
                410                 415                 420
Glu Glu Gln Arg Lys Gln Lys Lys Arg Gln Ser Val Ser Gly Leu
                425                 430                 435
His Arg Tyr Leu His Leu Leu Asp Gly Asn Glu Asn Tyr Pro Cys
                440                 445                 450
Leu Val Asp Ala Asp Gly Asp Val Ile Ser Phe Pro Pro Ile Thr
                455                 460                 465
Asn Ser Glu Lys Thr Lys Val Lys Lys Thr Thr Ser Asp Leu Phe
                470                 475                 480
Leu Glu Val Thr Ser Ala Thr Ser Leu Gln Ile Cys Lys Asp Val
                485                 490                 495
Met Asp Ala Leu Ile Leu Lys Met Ala Glu Met Lys Lys Tyr Thr
                500                 505                 510
Leu Glu Asn Lys Glu Glu Gly Ser Leu Ser Asp Thr Glu Ala Asp
                515                 520                 525
Ala Val Ser Gly Gln Leu Pro Asp Pro Thr Thr Asn Pro Ser Ala
                530                 535                 540
Gly Lys Asp Gly Pro Ser Leu Leu Val Val Glu Gln Val Arg Val
                545                 550                 555
Val Asp Leu Glu Gly Ser Leu Lys Val Val Tyr Pro Ser Lys Ala
                560                 565                 570
Asp Leu Ala Thr Ala Pro Pro His Val Thr Val Val Arg
                575                 580
```

<210> 14
<211> 309
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4650669CD1

<400> 14

```
Met Ser Asp Leu Gly Ser Glu Glu Leu Glu Glu Glu Gly Glu Asn
  1               5                  10                  15
Asp Ile Gly Glu Tyr Glu Gly Gly Arg Asn Glu Ala Gly Glu Arg
                 20                  25                  30
His Gly Arg Gly Arg Ala Arg Leu Pro Asn Gly Asp Thr Tyr Glu
                 35                  40                  45
Gly Ser Tyr Glu Phe Gly Lys Arg His Gly Gln Gly Ile Tyr Lys
                 50                  55                  60
Phe Lys Asn Gly Ala Arg Tyr Ile Gly Glu Tyr Val Arg Asn Lys
                 65                  70                  75
Lys His Gly Gln Gly Thr Phe Ile Tyr Pro Asp Gly Ser Arg Tyr
```

```
                 80                  85                  90
Glu Gly Glu Trp Ala Asn Asp Leu Arg His Gly His Gly Val Tyr
                 95                 100                 105
Tyr Tyr Ile Asn Asn Asp Thr Tyr Thr Gly Glu Trp Phe Ala His
                110                 115                 120
Gln Arg His Gly Gln Gly Thr Tyr Leu Tyr Ala Glu Thr Gly Ser
                125                 130                 135
Lys Tyr Val Gly Thr Trp Val Asn Gly Gln Gln Glu Gly Thr Ala
                140                 145                 150
Glu Leu Ile His Leu Asn His Arg Tyr Gln Gly Lys Phe Leu Asn
                155                 160                 165
Lys Asn Pro Val Gly Pro Gly Lys Tyr Val Phe Asp Val Gly Cys
                170                 175                 180
Glu Gln His Gly Glu Tyr Arg Leu Thr Asp Met Glu Arg Gly Glu
                185                 190                 195
Glu Glu Glu Glu Glu Glu Leu Val Thr Val Val Pro Lys Trp Lys
                200                 205                 210
Ala Thr Gln Ile Thr Glu Leu Ala Leu Trp Thr Pro Thr Leu Pro
                215                 220                 225
Lys Lys Pro Thr Ser Thr Asp Gly Pro Gly Gln Asp Ala Pro Gly
                230                 235                 240
Ala Glu Ser Ala Gly Glu Pro Gly Glu Glu Ala Gln Ala Leu Leu
                245                 250                 255
Glu Gly Phe Glu Gly Glu Met Asp Met Arg Pro Gly Asp Glu Asp
                260                 265                 270
Ala Asp Val Leu Arg Glu Glu Ser Arg Glu Tyr Asp Gln Glu Glu
                275                 280                 285
Phe Arg Tyr Asp Met Asp Glu Gly Asn Ile Asn Ser Glu Glu Glu
                290                 295                 300
Glu Thr Arg Gln Ser Asp Leu Gln Asp
                305
```

<210> 15
<211> 252
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485268CD1

<400> 15

```
Met Ala Ala Pro Ala Leu Leu Leu Leu Ala Leu Leu Leu Pro Val
  1               5                  10                  15
Gly Ala Trp Pro Gly Leu Pro Arg Arg Pro Cys Val His Cys Cys
                 20                  25                  30
Arg Pro Ala Trp Pro Pro Gly Pro Tyr Ala Arg Val Ser Asp Arg
                 35                  40                  45
Asp Leu Trp Arg Gly Asp Leu Trp Arg Gly Leu Pro Arg Val Arg
                 50                  55                  60
Pro Thr Ile Asp Ile Glu Ile Leu Lys Gly Glu Lys Gly Glu Ala
                 65                  70                  75
Gly Val Arg Gly Arg Ala Gly Arg Ser Gly Lys Glu Gly Pro Pro
                 80                  85                  90
Gly Ala Arg Gly Leu Gln Gly Arg Arg Gly Gln Lys Gly Gln Val
                 95                 100                 105
```

```
Gly Pro Pro Gly Ala Ala Cys Arg Arg Ala Tyr Ala Ala Phe Ser
                110                 115                 120
Val Gly Arg Arg Glu Gly Leu His Ser Ser Asp His Phe Gln Ala
                125                 130                 135
Val Pro Phe Asp Thr Glu Leu Val Asn Leu Asp Gly Ala Phe Asp
                140                 145                 150
Leu Ala Ala Gly Arg Phe Leu Cys Thr Val Pro Gly Val Tyr Phe
                155                 160                 165
Leu Ser Leu Asn Val His Thr Trp Asn Tyr Lys Glu Thr Tyr Leu
                170                 175                 180
His Ile Met Leu Asn Arg Arg Pro Ala Ala Val Leu Tyr Ala Gln
                185                 190                 195
Pro Ser Glu Arg Ser Val Met Gln Ala Gln Ser Leu Met Leu Leu
                200                 205                 210
Leu Ala Ala Gly Asp Ala Val Trp Val Arg Met Phe Gln Arg Asp
                215                 220                 225
Arg Asp Asn Ala Ile Tyr Gly Glu His Gly Asp Leu Tyr Ile Thr
                230                 235                 240
Phe Ser Gly His Leu Val Lys Pro Ala Ala Glu Leu
                245                 250
```

<210> 16
<211> 667
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2112995CD1

<400> 16

```
Met Ser Ser Gln Pro Ala Gly Asn Gln Thr Ser Pro Gly Ala Thr
  1               5                  10                  15
Glu Asp Tyr Ser Tyr Gly Ser Trp Tyr Ile Asp Glu Pro Gln Gly
                20                  25                  30
Gly Glu Glu Leu Gln Pro Glu Gly Glu Val Pro Ser Cys His Thr
                35                  40                  45
Ser Ile Pro Pro Gly Leu Tyr His Ala Cys Leu Ala Ser Leu Ser
                50                  55                  60
Ile Leu Val Leu Leu Leu Leu Ala Met Leu Val Arg Arg Arg Gln
                65                  70                  75
Leu Trp Pro Asp Cys Val Arg Gly Arg Pro Gly Leu Pro Ser Pro
                80                  85                  90
Val Asp Phe Leu Ala Gly Asp Arg Pro Arg Ala Val Pro Ala Ala
                95                 100                 105
Val Phe Met Val Leu Leu Ser Ser Leu Cys Leu Leu Leu Pro Asp
                110                 115                 120
Glu Asp Ala Leu Pro Phe Leu Thr Leu Ala Ser Ala Pro Ser Gln
                125                 130                 135
Asp Gly Lys Thr Glu Ala Pro Arg Gly Ala Trp Lys Ile Leu Gly
                140                 145                 150
Leu Phe Tyr Tyr Ala Ala Leu Tyr Tyr Pro Leu Ala Ala Cys Ala
                155                 160                 165
Thr Ala Gly His Thr Ala Ala His Leu Leu Gly Ser Thr Leu Ser
                170                 175                 180
Trp Ala His Leu Gly Val Gln Val Trp Gln Arg Ala Glu Cys Pro
```

```
                185                 190                 195
Gln Val Pro Lys Ile Tyr Lys Tyr Tyr Ser Leu Leu Ala Ser Leu
                200                 205                 210
Pro Leu Leu Leu Gly Leu Gly Phe Leu Ser Leu Trp Tyr Pro Val
                215                 220                 225
Gln Leu Val Arg Ser Phe Ser Arg Arg Thr Gly Ala Gly Ser Lys
                230                 235                 240
Gly Leu Gln Ser Ser Tyr Ser Glu Glu Tyr Leu Arg Asn Leu Leu
                245                 250                 255
Cys Arg Lys Lys Leu Gly Ser Ser Tyr His Thr Ser Lys His Gly
                260                 265                 270
Phe Leu Ser Trp Ala Arg Val Cys Leu Arg His Cys Ile Tyr Thr
                275                 280                 285
Pro Gln Pro Gly Phe His Leu Pro Leu Lys Leu Val Leu Ser Ala
                290                 295                 300
Thr Leu Thr Gly Thr Ala Ile Tyr Gln Val Ala Leu Leu Leu Leu
                305                 310                 315
Val Gly Val Val Pro Thr Ile Gln Lys Val Arg Ala Gly Val Thr
                320                 325                 330
Thr Asp Val Ser Tyr Leu Leu Ala Gly Phe Gly Ile Val Leu Ser
                335                 340                 345
Glu Asp Lys Gln Glu Val Val Glu Leu Val Lys His His Leu Trp
                350                 355                 360
Ala Leu Glu Val Cys Tyr Ile Ser Ala Leu Val Leu Ser Cys Leu
                365                 370                 375
Leu Thr Phe Leu Val Leu Met Arg Ser Leu Val Thr His Arg Thr
                380                 385                 390
Asn Leu Arg Ala Leu His Arg Gly Ala Ala Leu Asp Leu Ser Pro
                395                 400                 405
Leu His Arg Ser Pro His Pro Ser Arg Gln Ala Ile Phe Cys Trp
                410                 415                 420
Met Ser Phe Ser Ala Tyr Gln Thr Ala Phe Ile Cys Leu Gly Leu
                425                 430                 435
Leu Val Gln Gln Ile Ile Phe Phe Leu Gly Thr Thr Ala Leu Ala
                440                 445                 450
Phe Leu Val Leu Met Pro Val Leu His Gly Arg Asn Leu Leu Leu
                455                 460                 465
Phe Arg Ser Leu Glu Ser Ser Trp Pro Phe Trp Leu Thr Leu Ala
                470                 475                 480
Leu Ala Val Ile Leu Gln Asn Met Ala Ala His Trp Val Phe Leu
                485                 490                 495
Glu Thr His Asp Gly His Pro Gln Leu Thr Asn Arg Arg Val Leu
                500                 505                 510
Tyr Ala Ala Thr Phe Leu Leu Phe Pro Leu Asn Val Leu Val Gly
                515                 520                 525
Ala Met Val Ala Thr Trp Arg Val Leu Leu Ser Ala Leu Tyr Asn
                530                 535                 540
Ala Ile His Leu Gly Gln Met Asp Leu Ser Leu Leu Pro Pro Arg
                545                 550                 555
Ala Ala Thr Leu Asp Pro Gly Tyr Tyr Thr Tyr Arg Asn Phe Leu
                560                 565                 570
Lys Ile Glu Val Ser Gln Ser His Pro Ala Met Thr Ala Phe Cys
                575                 580                 585
Ser Leu Leu Leu Gln Ala Gln Ser Leu Leu Pro Arg Thr Met Ala
                590                 595                 600
Ala Pro Gln Asp Ser Leu Arg Pro Gly Glu Glu Asp Glu Gly Met
```

```
                605                 610                 615
Gln Leu Leu Gln Thr Lys Asp Ser Met Ala Lys Gly Ala Arg Pro
                620                 625                 630
Gly Ala Ser Arg Gly Arg Ala Arg Trp Gly Leu Ala Tyr Thr Leu
                635                 640                 645
Leu His Asn Pro Thr Leu Gln Val Phe Arg Lys Thr Ala Leu Leu
                650                 655                 660
Gly Ala Asn Gly Ala Gln Pro
                665
```

<210> 17
<211> 657
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1613452CD1

<400> 17

```
Met Ala Glu Gly Ser Gly Glu Val Val Thr Val Ser Ala Thr Gly
  1                 5                  10                  15
Ala Ala Asn Gly Leu Asn Asn Gly Ala Gly Gly Thr Ser Ala Thr
                 20                  25                  30
Thr Ser Asn Pro Leu Ser Arg Lys Leu His Lys Ile Leu Glu Thr
                 35                  40                  45
Arg Leu Asp Asn Asp Lys Glu Met Leu Glu Ala Leu Lys Ala Leu
                 50                  55                  60
Ser Thr Phe Phe Val Glu Asn Ser Leu Arg Thr Arg Arg Asn Leu
                 65                  70                  75
Arg Gly Asp Ile Glu Arg Lys Ser Leu Ala Ile Asn Glu Glu Phe
                 80                  85                  90
Val Ser Ile Phe Lys Glu Val Lys Glu Glu Leu Glu Ser Ile Ser
                 95                 100                 105
Glu Asp Val Gln Ala Met Ser Asn Cys Cys Gln Asp Met Thr Ser
                110                 115                 120
Arg Leu Gln Ala Ala Lys Glu Gln Thr Gln Asp Leu Ile Val Lys
                125                 130                 135
Thr Thr Lys Leu Gln Ser Glu Ser Gln Lys Leu Glu Ile Arg Ala
                140                 145                 150
Gln Val Ala Asp Ala Phe Leu Ser Lys Phe Gln Leu Thr Ser Asp
                155                 160                 165
Glu Met Ser Leu Leu Arg Gly Thr Arg Glu Gly Pro Ile Thr Glu
                170                 175                 180
Asp Phe Phe Lys Ala Leu Gly Arg Val Lys Gln Ile His Asn Asp
                185                 190                 195
Val Lys Val Leu Leu Arg Thr Asn Gln Gln Thr Ala Gly Leu Glu
                200                 205                 210
Ile Met Glu Gln Met Ala Leu Leu Gln Glu Thr Ala Tyr Glu Arg
                215                 220                 225
Leu Tyr Arg Trp Ala Gln Ser Glu Cys Arg Thr Leu Thr Gln Glu
                230                 235                 240
Ser Cys Asp Val Ser Pro Val Leu Thr Gln Ala Met Glu Ala Leu
                245                 250                 255
Gln Asp Arg Pro Val Leu Tyr Lys Tyr Thr Leu Asp Glu Phe Gly
                260                 265                 270
```

```
Thr Ala Arg Arg Ser Thr Val Val Arg Gly Phe Ile Asp Ala Leu
                275                 280                 285
Thr Arg Gly Gly Pro Gly Gly Thr Pro Arg Pro Ile Glu Met His
                290                 295                 300
Ser His Asp Pro Leu Arg Tyr Val Gly Asp Met Leu Ala Trp Leu
                305                 310                 315
His Gln Ala Thr Ala Ser Glu Lys Glu His Leu Glu Ala Leu Leu
                320                 325                 330
Lys His Val Thr Thr Gln Gly Val Glu Glu Asn Ile Gln Glu Val
                335                 340                 345
Val Gly His Ile Thr Glu Gly Val Cys Arg Pro Leu Lys Val Arg
                350                 355                 360
Ile Glu Gln Val Ile Val Ala Glu Pro Gly Ala Val Leu Leu Tyr
                365                 370                 375
Lys Ile Ser Asn Leu Leu Lys Phe Tyr His His Thr Ile Ser Gly
                380                 385                 390
Ile Val Gly Asn Ser Ala Thr Ala Leu Leu Thr Thr Ile Glu Glu
                395                 400                 405
Met His Leu Leu Ser Lys Lys Ile Phe Phe Asn Ser Leu Ser Leu
                410                 415                 420
His Ala Ser Lys Leu Met Asp Lys Val Glu Leu Pro Pro Pro Asp
                425                 430                 435
Leu Gly Pro Ser Ser Ala Leu Asn Gln Thr Leu Met Leu Leu Arg
                440                 445                 450
Glu Val Leu Ala Ser His Asp Ser Ser Val Val Pro Leu Asp Ala
                455                 460                 465
Arg Gln Ala Asp Phe Val Gln Val Leu Ser Cys Val Leu Asp Pro
                470                 475                 480
Leu Leu Gln Met Cys Thr Val Ser Ala Ser Asn Leu Gly Thr Ala
                485                 490                 495
Asp Met Ala Thr Phe Met Val Asn Ser Leu Tyr Met Met Lys Thr
                500                 505                 510
Thr Leu Ala Leu Phe Glu Phe Thr Asp Arg Arg Leu Glu Met Leu
                515                 520                 525
Gln Phe Gln Ile Glu Ala His Leu Asp Thr Leu Ile Asn Glu Gln
                530                 535                 540
Ala Ser Tyr Val Leu Thr Arg Val Gly Leu Ser Tyr Ile Tyr Asn
                545                 550                 555
Thr Val Gln Gln His Lys Pro Glu Gln Gly Ser Leu Ala Asn Met
                560                 565                 570
Pro Asn Leu Asp Ser Val Thr Leu Lys Ala Ala Met Val Gln Phe
                575                 580                 585
Asp Arg Tyr Leu Ser Ala Pro Asp Asn Leu Leu Ile Pro Gln Leu
                590                 595                 600
Asn Phe Leu Leu Ser Ala Thr Val Lys Glu Gln Ile Val Lys Gln
                605                 610                 615
Ser Thr Glu Leu Val Cys Arg Ala Tyr Gly Glu Val Tyr Ala Ala
                620                 625                 630
Val Met Asn Pro Ile Asn Glu Tyr Lys Asp Pro Glu Asn Ile Leu
                635                 640                 645
His Arg Ser Pro Gln Gln Val Gln Thr Leu Leu Ser
                650                 655
```

<210> 18
<211> 1958
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55061615CD1

<400> 18

```
Met Thr Ile Leu Asn Ala Leu Leu His Ala Asp Pro Val Gln Gly
  1               5                  10                  15
Lys Arg Ile Gln Leu Lys Ala Arg Ala Phe Glu Leu Ser Glu Gly
                 20                  25                  30
Asp Val Leu Lys Val Tyr Asp Gly Asn Asn Asn Ser Ala Arg Leu
                 35                  40                  45
Leu Gly Val Phe Ser His Ser Glu Met Met Gly Val Thr Leu Asn
                 50                  55                  60
Ser Thr Ser Ser Ser Leu Trp Leu Asp Phe Ile Thr Asp Ala Glu
                 65                  70                  75
Asn Thr Ser Lys Gly Phe Glu Leu His Phe Ser Ser Phe Glu Leu
                 80                  85                  90
Ile Lys Cys Glu Asp Pro Gly Thr Pro Lys Phe Gly Tyr Lys Val
                 95                 100                 105
His Asp Glu Gly His Phe Ala Gly Ser Ser Val Ser Phe Ser Cys
                110                 115                 120
Asp Pro Gly Tyr Ser Leu Arg Gly Ser Glu Glu Leu Leu Cys Leu
                125                 130                 135
Ser Gly Glu Arg Arg Thr Trp Asp Arg Pro Leu Pro Thr Cys Val
                140                 145                 150
Ala Glu Cys Gly Gly Thr Val Arg Gly Glu Val Ser Gly Gln Val
                155                 160                 165
Leu Ser Pro Gly Tyr Pro Ala Pro Tyr Glu His Asn Leu Asn Cys
                170                 175                 180
Ile Trp Thr Ile Glu Ala Glu Ala Gly Cys Thr Ile Gly Leu His
                185                 190                 195
Phe Leu Val Phe Asp Thr Glu Glu Val His Asp Val Leu Arg Ile
                200                 205                 210
Trp Asp Gly Pro Val Glu Ser Gly Val Leu Leu Lys Glu Leu Ser
                215                 220                 225
Gly Pro Ala Leu Pro Lys Asp Leu His Ser Thr Phe Asn Ser Val
                230                 235                 240
Val Leu Gln Phe Ser Thr Asp Phe Phe Thr Ser Lys Gln Gly Phe
                245                 250                 255
Ala Ile Gln Phe Ser Val Ser Thr Ala Thr Ser Cys Asn Asp Pro
                260                 265                 270
Gly Ile Pro Gln Asn Gly Ser Arg Ser Gly Asp Ser Trp Glu Ala
                275                 280                 285
Gly Asp Ser Thr Val Phe Gln Cys Asp Pro Gly Tyr Ala Leu Gln
                290                 295                 300
Gly Ser Ala Glu Ile Ser Cys Val Lys Ile Glu Asn Arg Phe Phe
                305                 310                 315
Trp Gln Pro Ser Pro Pro Thr Cys Ile Ala Pro Cys Gly Gly Asp
                320                 325                 330
Leu Thr Gly Pro Ser Gly Val Ile Leu Ser Pro Asn Tyr Pro Glu
                335                 340                 345
Pro Tyr Pro Pro Gly Lys Glu Cys Asp Trp Lys Val Thr Val Ser
                350                 355                 360
Pro Asp Tyr Val Ile Ala Leu Val Phe Asn Ile Phe Asn Leu Glu
```

```
                365                 370                 375
Pro Gly Tyr Asp Phe Leu His Ile Tyr Asp Gly Arg Asp Ser Leu
                380                 385                 390
Ser Pro Leu Ile Gly Ser Phe Tyr Gly Ser Gln Leu Pro Gly Arg
                395                 400                 405
Ile Glu Ser Ser Ser Asn Ser Leu Phe Leu Ala Phe Arg Ser Asp
                410                 415                 420
Ala Ser Val Ser Asn Ala Gly Phe Val Ile Asp Tyr Thr Glu Asn
                425                 430                 435
Pro Arg Glu Ser Cys Phe Asp Pro Gly Ser Ile Lys Ser Gly Thr
                440                 445                 450
Arg Val Gly Ser Asp Leu Lys Leu Gly Ser Ser Val Thr Tyr Tyr
                455                 460                 465
Cys His Gly Gly Tyr Glu Val Glu Gly Thr Ser Thr Leu Ser Cys
                470                 475                 480
Ile Leu Gly Pro Asp Gly Lys Pro Val Trp Asn Asn Pro Arg Pro
                485                 490                 495
Val Cys Thr Ala Pro Cys Gly Gly Gln Tyr Val Gly Ser Asp Gly
                500                 505                 510
Val Val Leu Ser Pro Asn Tyr Pro Gln Asn Tyr Thr Ser Gly Gln
                515                 520                 525
Ile Cys Leu Tyr Phe Val Thr Val Pro Lys Asp Tyr Val Val Phe
                530                 535                 540
Gly Gln Phe Ala Phe Phe His Thr Ala Leu Asn Asp Val Val Glu
                545                 550                 555
Val His Asp Gly His Ser Gln His Ser Arg Leu Leu Ser Ser Leu
                560                 565                 570
Ser Gly Ser His Thr Gly Ile Arg Gly Ser Ala Ser Val Gly Met
                575                 580                 585
Val Val Gly Arg Gly His His Val Arg Leu Lys Glu Gly Gly Ser
                590                 595                 600
Arg Ser Thr Pro Trp Pro Gln Val Glu Pro Tyr Gly Ser Ala Cys
                605                 610                 615
Leu Ser Cys Ser Gly Ala Cys Leu Gln Arg Ser Ser Gln Leu Val
                620                 625                 630
Arg Ala Pro Thr Ser Gly Ala Phe Ser Ser Cys Pro His Pro Asp
                635                 640                 645
Cys Val Tyr Thr Ala Pro Leu Trp Cys Ser Leu Leu Leu Leu Asn
                650                 655                 660
Gly Asn Tyr Thr Asn Trp Leu Gln Val Gln Leu Val Leu Ser Leu
                665                 670                 675
Pro Trp Pro Ile Cys Thr Ala Pro Ser Arg Arg Tyr Thr Phe Val
                680                 685                 690
Phe Cys Tyr Lys Ser Cys Gln Ser Thr Leu Val Ser Cys Ala His
                695                 700                 705
Ala Gly Glu Ser Leu Pro Leu Ala Thr Ser Asn Gln Val Leu Ile
                710                 715                 720
Lys Phe Ser Ala Lys Gly Leu Ala Pro Ala Arg Gly Phe His Phe
                725                 730                 735
Val Tyr Gln Gly Met Glu Asp Met Asp Ala Gly Ala Val Pro Arg
                740                 745                 750
Thr Ser Ala Thr Gln Cys Ser Ser Val Pro Glu Pro Arg Tyr Gly
                755                 760                 765
Lys Arg Leu Gly Ser Asp Phe Ser Val Gly Ala Ile Val Arg Phe
                770                 775                 780
Glu Cys Asn Ser Gly Tyr Ala Leu Gln Gly Ser Pro Glu Ile Glu
```

```
                785                 790                 795
Cys Leu Pro Val Pro Gly Ala Leu Ala Gln Trp Asn Val Ser Ala
                800                 805                 810
Pro Thr Cys Val Val Pro Cys Gly Gly Asn Leu Thr Glu Arg Arg
                815                 820                 825
Gly Thr Ile Leu Ser Pro Gly Phe Pro Glu Pro Tyr Leu Asn Ser
                830                 835                 840
Leu Asn Cys Val Trp Lys Ile Val Val Pro Glu Gly Ala Gly Ile
                845                 850                 855
Gln Ile Gln Val Val Ser Phe Val Thr Glu Gln Asn Trp Asp Ser
                860                 865                 870
Leu Glu Val Phe Asp Gly Ala Asp Asn Thr Val Thr Met Leu Gly
                875                 880                 885
Ser Phe Ser Gly Thr Thr Val Pro Ala Leu Leu Asn Ser Thr Ser
                890                 895                 900
Asn Gln Leu Tyr Leu His Phe Tyr Ser Asp Ile Ser Val Ser Ala
                905                 910                 915
Ala Gly Phe His Leu Glu Tyr Lys Thr Val Gly Leu Ser Ser Cys
                920                 925                 930
Pro Glu Pro Ala Val Pro Ser Asn Gly Val Lys Thr Gly Glu Arg
                935                 940                 945
Tyr Leu Val Asn Asp Val Val Ser Phe Gln Cys Glu Pro Gly Tyr
                950                 955                 960
Ala Leu Gln Gly His Ala His Ile Ser Cys Met Pro Gly Thr Val
                965                 970                 975
Arg Arg Trp Asn Tyr Pro Pro Pro Leu Cys Ile Ala Gln Cys Gly
                980                 985                 990
Gly Thr Val Glu Glu Met Glu Gly Val Ile Leu Ser Pro Gly Phe
                995                1000                1005
Pro Gly Asn Tyr Pro Ser Asn Met Asp Cys Ser Trp Lys Ile Ala
               1010                1015                1020
Leu Pro Val Gly Phe Gly Ala His Ile Gln Phe Leu Asn Phe Ser
               1025                1030                1035
Thr Glu Pro Asn His Asp Tyr Ile Glu Ile Arg Asn Gly Pro Tyr
               1040                1045                1050
Glu Thr Ser Arg Met Met Gly Arg Phe Ser Gly Ser Glu Leu Pro
               1055                1060                1065
Ser Ser Leu Leu Ser Thr Ser His Glu Thr Thr Val Tyr Phe His
               1070                1075                1080
Ser Asp His Ser Gln Asn Arg Pro Gly Phe Lys Leu Glu Tyr Gln
               1085                1090                1095
Asp Leu Thr Tyr Ser His Gln Ile Ser Ser Phe Leu Arg Gly Phe
               1100                1105                1110
Asp Leu Ser Glu Leu Glu Arg Thr Asn Ser Thr Pro Pro Val Ala
               1115                1120                1125
Ala Ser Tyr Val Trp Asp Leu Asp Pro Gly Cys Glu Ala Tyr Glu
               1130                1135                1140
Leu Gln Glu Cys Pro Asp Pro Glu Pro Phe Ala Asn Gly Ile Val
               1145                1150                1155
Arg Gly Ala Gly Tyr Asn Val Gly Gln Ser Val Thr Phe Glu Cys
               1160                1165                1170
Leu Pro Gly Tyr Gln Leu Thr Gly His Pro Val Leu Thr Cys Gln
               1175                1180                1185
His Gly Thr Asn Arg Asn Trp Asp His Pro Leu Pro Lys Cys Glu
               1190                1195                1200
Val Pro Cys Gly Gly Asn Ile Thr Ser Ser Asn Gly Thr Val Tyr
```

```
                1205                1210                1215
Ser Pro Gly Phe Pro Ser Pro Tyr Ser Ser Ser Gln Asp Cys Val
                1220                1225                1230
Trp Leu Ile Thr Val Ala Gln Leu Ala Met Gly Val Arg Leu Asn
                1235                1240                1245
Leu Ser Leu Leu Gln Thr Glu Pro Ser Gly Asp Phe Ile Thr Ile
                1250                1255                1260
Trp Asp Gly Pro Gln Gln Thr Ala Pro Arg Leu Gly Val Phe Thr
                1265                1270                1275
Arg Ser Met Ala Lys Lys Thr Val Gln Ser Ser Ser Asn Gln Val
                1280                1285                1290
Leu Leu Lys Phe His Arg Asp Ala Ala Thr Gly Gly Ile Phe Ala
                1295                1300                1305
Ile Ala Phe Ser Ala Tyr Pro Leu Thr Lys Cys Pro Pro Pro Thr
                1310                1315                1320
Ile Leu Pro Asn Ala Glu Val Val Thr Glu Asn Glu Glu Phe Asn
                1325                1330                1335
Ile Gly Asp Ile Val Arg Tyr Arg Cys Leu Pro Gly Phe Thr Leu
                1340                1345                1350
Val Gly Asn Glu Ile Leu Thr Cys Lys Leu Gly Thr Tyr Leu Gln
                1355                1360                1365
Phe Glu Gly Pro Pro Pro Ile Cys Glu Val His Cys Pro Thr Asn
                1370                1375                1380
Glu Leu Leu Thr Asp Ser Thr Gly Val Ile Leu Ser Gln Ser Tyr
                1385                1390                1395
Pro Gly Ser Tyr Pro Gln Phe Gln Thr Cys Ser Trp Leu Val Arg
                1400                1405                1410
Val Glu Pro Asp Tyr Asn Ile Ser Leu Thr Val Glu Tyr Phe Leu
                1415                1420                1425
Ser Glu Lys Gln Tyr Asp Glu Phe Glu Ile Phe Asp Gly Pro Ser
                1430                1435                1440
Gly Gln Ser Pro Leu Leu Lys Ala Leu Ser Gly Asn Tyr Ser Ala
                1445                1450                1455
Pro Leu Ile Val Thr Ser Ser Ser Asn Ser Val Tyr Leu Arg Trp
                1460                1465                1470
Ser Ser Asp His Ala Tyr Asn Arg Lys Gly Phe Lys Ile Arg Tyr
                1475                1480                1485
Ser Gly Gln Thr Ser Thr Gln Pro Gly Gly Ser Ile His Phe Gly
                1490                1495                1500
Cys Asn Ala Gly Tyr Arg Leu Val Gly His Ser Met Ala Ile Cys
                1505                1510                1515
Thr Arg His Pro Gln Gly Tyr His Leu Trp Ser Glu Ala Ile Pro
                1520                1525                1530
Leu Cys Gln Ala Leu Ser Cys Gly Leu Pro Glu Ala Pro Lys Asn
                1535                1540                1545
Gly Met Val Phe Gly Lys Glu Tyr Thr Val Gly Thr Lys Ala Met
                1550                1555                1560
Tyr Ser Cys Ser Glu Gly Tyr His Leu Gln Ala Gly Ala Glu Ala
                1565                1570                1575
Thr Ala Glu Cys Leu Asp Thr Gly Leu Trp Ser Asn Arg Asn Val
                1580                1585                1590
Pro Pro Gln Cys Val Arg Glu Ser Ser Gly Asn Gly Gly Gly Ser
                1595                1600                1605
Val Thr Cys Pro Asp Val Ser Ser Ile Ser Val Glu His Gly Arg
                1610                1615                1620
Trp Arg Leu Ile Phe Glu Thr Gln Tyr Gln Phe Gln Ala Gln Leu
```

```
                1625                1630                1635
Met Leu Ile Cys Asp Pro Gly Tyr Tyr Tyr Thr Gly Gln Arg Val
                1640                1645                1650
Ile Arg Cys Gln Ala Asn Gly Lys Trp Ser Leu Gly Asp Ser Thr
                1655                1660                1665
Pro Thr Cys Arg Ile Ile Ser Cys Gly Glu Leu Pro Ile Pro Pro
                1670                1675                1680
Asn Gly His Arg Ile Gly Thr Leu Ser Val Tyr Gly Ala Thr Ala
                1685                1690                1695
Ile Phe Ser Cys Asn Ser Gly Tyr Thr Leu Val Gly Ser Arg Val
                1700                1705                1710
Arg Glu Cys Met Ala Asn Gly Leu Trp Ser Gly Ser Glu Val Arg
                1715                1720                1725
Cys Leu Ala Thr Gln Thr Lys Leu His Ser Ile Phe Tyr Lys Leu
                1730                1735                1740
Leu Phe Asp Val Leu Ser Ser Pro Ser Leu Thr Lys Ala Gly His
                1745                1750                1755
Cys Gly Thr Pro Glu Pro Ile Val Asn Gly His Ile Asn Gly Glu
                1760                1765                1770
Asn Tyr Ser Tyr Arg Gly Ser Val Val Tyr Gln Cys Asn Ala Gly
                1775                1780                1785
Phe Arg Leu Ile Gly Met Ser Val Arg Ile Cys Gln Gln Asp His
                1790                1795                1800
His Trp Ser Gly Lys Thr Pro Phe Cys Val His Val Lys Gln Gln
                1805                1810                1815
Leu Leu Leu Leu Leu Leu Leu Leu Cys Asp Asp Asp Asp Asp Glu
                1820                1825                1830
Asp Asp Gly Ser Gly Ala Ile Thr Cys Gly His Pro Gly Asn Pro
                1835                1840                1845
Val Asn Gly Leu Thr Gln Gly Asn Gln Phe Asn Leu Asn Asp Val
                1850                1855                1860
Val Lys Phe Val Cys Asn Pro Gly Tyr Met Ala Glu Gly Ala Ala
                1865                1870                1875
Arg Ser Gln Cys Leu Ala Ser Gly Gln Trp Ser Asp Met Leu Pro
                1880                1885                1890
Thr Cys Arg Ile Ile Asn Cys Thr Asp Pro Gly His Gln Glu Asn
                1895                1900                1905
Ser Val Arg Gln Val His Ala Ser Gly Pro His Arg Phe Ser Phe
                1910                1915                1920
Gly Thr Thr Val Ser Tyr Arg Cys Asn His Gly Phe Tyr Leu Leu
                1925                1930                1935
Gly Thr Pro Val Leu Ser Cys Gln Gly Asp Gly Thr Trp Asp Arg
                1940                1945                1950
Pro Arg Pro Gln Cys Leu Cys Lys
                1955
```

<210> 19
<211> 100
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503435CD1

<400> 19

```
Met Lys Ser Lys Gly Val Lys Ser Tyr Gln Arg Arg Pro Arg Glu
  1               5                  10                  15
Glu Arg Thr Gln Arg Arg Thr Arg Cys Gln Ser Arg Arg Gly Ser
                20                  25                  30
Trp Arg Ser Arg His Trp Arg Trp Trp Asn Lys Leu Leu Pro Thr
                35                  40                  45
Pro Trp Met Thr Gly Thr Leu Gly Ser Ser Ser Cys Gln Leu Ser
                50                  55                  60
Cys Ala His Gln Pro Gly Thr Ala Gly Ile Trp Ala Glu Ala Leu
                65                  70                  75
Thr Arg Gln Cys Pro Gly Pro Gln Thr Ser Pro Pro Thr Ser Gln
                80                  85                  90
Asn Ile Pro Ser Glu Pro Gly Ser Phe Thr
                95                 100
```

<210> 20
<211> 271
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7504149CD1

<400> 20

```
Met Ser Asp Leu Gly Ser Glu Glu Leu Glu Glu Glu Gly Glu Asn
  1               5                  10                  15
Asp Ile Gly Gly Ile Tyr Lys Phe Lys Asn Gly Ala Arg Tyr Ile
                20                  25                  30
Gly Glu Tyr Val Arg Asn Lys Lys His Gly Gln Gly Thr Phe Ile
                35                  40                  45
Tyr Pro Asp Gly Ser Arg Tyr Glu Gly Glu Trp Ala Asn Asp Leu
                50                  55                  60
Arg His Gly His Gly Val Tyr Tyr Tyr Ile Asn Asn Asp Thr Tyr
                65                  70                  75
Thr Gly Glu Trp Phe Ala His Gln Arg His Gly Gln Gly Thr Tyr
                80                  85                  90
Leu Tyr Ala Glu Thr Gly Ser Lys Tyr Val Gly Thr Trp Val Asn
                95                 100                 105
Gly Gln Gln Glu Gly Thr Ala Glu Leu Ile His Leu Asn His Arg
               110                 115                 120
Tyr Gln Gly Lys Phe Leu Asn Lys Asn Pro Val Gly Pro Gly Lys
               125                 130                 135
Tyr Val Phe Asp Val Gly Cys Glu Gln His Gly Glu Tyr Arg Leu
               140                 145                 150
Thr Asp Met Glu Arg Gly Glu Glu Glu Glu Glu Glu Glu Leu Val
               155                 160                 165
Thr Val Val Pro Lys Trp Lys Ala Thr Gln Ile Thr Glu Leu Ala
               170                 175                 180
Leu Trp Thr Pro Thr Leu Pro Lys Lys Pro Thr Ser Thr Asp Gly
               185                 190                 195
Pro Gly Gln Asp Ala Pro Gly Ala Glu Ser Ala Gly Glu Pro Gly
               200                 205                 210
Glu Glu Ala Gln Ala Leu Leu Glu Gly Phe Glu Gly Glu Met Asp
               215                 220                 225
Met Arg Pro Gly Asp Glu Asp Ala Asp Val Leu Arg Glu Glu Ser
```

```
                230                 235                 240
Arg Glu Tyr Asp Gln Glu Glu Phe Arg Tyr Asp Met Asp Glu Gly
                245                 250                 255
Asn Ile Asn Ser Glu Glu Glu Glu Thr Arg Gln Ser Asp Leu Gln
                260                 265                 270
Asp
```

<210> 21
<211> 1506
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1419725CB1

<400> 21
ctcaaaggca aacaaaagga aatgcccggc tccccatggc tgtggccagc accttcatac 60
cagggctcaa ccctcagaac cctcattata tcccagggta agactcccac ttagcgctcc 120
ccgcttgctg ctgggagtag gggtataggg tggaggaatt aagctgctta gagcatgggg 180
ccaagcggac aggttggttt gggaggctga ggtctctatg gtggtggatg gaaaaggaga 240
agcaagggac tagtgagaca ccgggtctgc cccaaagctt tgtgttgttc tctccatggg 300
aacagggtaa cagcagagga aggatcgaaa gcccaggcag agaacttaga aatgcctgct 360
ccaaatatat acactgctga gtccagaaag atcagggctt cagaggtctc cgcttctgcc 420
tctggggatg gcggggcagg gcgattctgg tgagtgggag agtctgtgct gcaggtacac 480
tggacactgc ccactacttc ggttcagcgt gggccagacc tatgggcagg tgactggtca 540
gctacttcga ggccctcctg gcctagcctg gccccctgtc caccgcacac ttctgcctcc 600
cattcggcct ccaagatctc ctgaggttcc cagggagagt ctacctgtca ggcgtgggca 660
agaaaggctc agctccagca tgatccctgg gtacacaggt ttgtacgcag gtatgcacag 720
gtgcccctcc caggagcatg tgttccaaac actaacgagt cttccttgtc cctgcctgcc 780
caggttttgt accccgggca cagttcatct ttgccaagaa ctgcagccag gtctgggccg 840
aggctctgag tgactttact cacttgcatg aaaagcaagg gagtgaagag ctaccaaagg 900
aggccaaggg aagaaaggac acagagaagg accaggtgcc agagccggag gggcagctgg 960
aggagccgac actggaggtg gtggaacaag cttctcccta ctccatggat gacagggacc 1020
ctcggaagtt cttcatgtca ggcttcactg gctatgtgcc ctgcgcccgc ttcctcttcg 1080
gctccagctt tcctgtgctc accaaccagg cactgcagga atttgggcag aagcactcac 1140
caggcagtgc ccaggacccc aaacatctcc ccccacttcc cagaacatac cctcagaacc 1200
tgggtctttt acctaactat gggggctacg tgccagggta taagttccag tttggccaca 1260
catttggcca tctcacccat gatgctctgg gcctcagcac cttccagaag cagctcttgg 1320
cttaggccac tggacatcaa gttcccttcc cttttcatcc tatcccagcc atccttttgg 1380
aagggagaga ggtgggtggg agggtgggag ggtgggggaa cacaaagaga aaatggtttg 1440
gaggctgagc acctttttta ttaataggta taataaataa ataaataaat acataaacag 1500
aaaaaa 1506

<210> 22
<211> 1565
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 628613CB1

<400> 22
ggctggttac gggtcctggc ccgcggagtt tgatttcgtc ctgggatgcg gaagtagcag 60

tcctgtcatg cggaagtagc agtccggtcc tagggactag caggcaccaa gaaactgata 120
atgttccttt gaattggctt ctgtatttgc ttcatcaatg tctctcatac tgaatatctt 180
aagagagatg ctggaatatt ttggcgttcc tgtagaacag gttttgctga tttgggaaaa 240
taaagactat ggatcaacta ggagtattgt tcgtattatt gggaaaatgc ttccactgga 300
accttgtcga agacctaatt ttgagttgat cccgctcttg aactctgtag actctgataa 360
ttgtggatct atggttccat cttttgctga tattttgtat gtggcaaatg atgaagaagc 420
cagttatctc agatttcgaa atagtatatg gaaaaatgaa gaagagaaag tggaaatttt 480
tcatcctttg cgactagttc gggatccact gtcacctgct gtaagacaga aagaaactgt 540
gaaaaatgac ctgcctgtaa atgaagctgc aattagaaaa atagctgccc ttgaaaatga 600
gctgactttt cttcgctctc agattgcagc aattgtggaa atgcaggaac tgaaaaatag 660
tacaaattct agttcctttg gcttgagtga cgagcgcatt agtttgggtc agctgtcatc 720
atcgcgggct gcccatctga gtgtggaccc agatcagctt ccaggttcag tgctttctcc 780
tcctcctcct ccaccacttc ctcctcagtt ttcatctctc cagccaccgt gttttcctcc 840
cgtacaacca ggatctaata atatttgtga ctcagataat ccagcaactg aaatgagcaa 900
acagaacccg gctgctaata agaccaatta tagtcatcat tcaaaaagcc agagaaataa 960
agatattcca aacatgttgg acgttctaaa ggatatgaat aaggttaagc ttcgtgcaat 1020
tgagcggtca cctggcggta gacccattca taagaggaaa agacagaatt cacattggga 1080
tccagtttct ttaatatctc atgcacttaa acagaaattt gcatttcaag aagatgattc 1140
ttttgagaaa gagaatagat cttgggaatc ttccccattt tctagtccag aaacttcaag 1200
gtttggacat cacatttcac agtcagaagg acagcgaact aaagaagaaa tggtcaacac 1260
aaaagctgtt gaccaaggta tcagcaacac aagccttcta aactcaagga tttaaactca 1320
acttaaggtt gagctttaaa cttccaaaac ttcttcctgg atgataaatt attcttagaa 1380
actgatttgg actgttaaag gctaaaagta gatgtattta aagactcttc ttgacacatt 1440
ttgcctacac ttgctatgta aatatgtatg cctgtcattt ttgtttcctt tgttcctttt 1500
tacgtttata ctctgttctt ctgtacatag agcttaaaat aaacattctt tttgaacttg 1560
aaaaa 1565

<210> 23
<211> 2488
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7111920CB1

<400> 23
ggcggcggcc gaggcggcgt cgttatttcc gtggtccgga cagtgcgtgg cggcgcgggt 60
gaccacggga gaagtaggca taatggttat gaaagcttct gtagatgatg acgattcagg 120
atgggagctc agtatgccag aaaaaatgga gaaaagcaat acaaactggg tggacattac 180
ccaagatttt gaagaagctt gtcgagaatt aaagttggga gaactacttc atgataagct 240
atttggtctt tttgaagcca tgtctgctat tgaaatgatg gatcccaaga tggatgctgg 300
catgattgga aaccaagtta atcgaaaagt tctcaatttt gaacaagcta tcaaggatgg 360
cactattaaa attaaagatc tcaccttgcc tgaactgata gggattatgg atacatgttt 420
ttgctgtttg ataacgtggt tagaaggcca ttcactggca cagacagtat ttacgtgcct 480
ttacattcat aatccagact ttatagaaga tcctgctatg aaggcttttg ctctgggaat 540
cttgaaaatc tgtgacattg caagggaaaa agtaaataaa gctgctgttt ttgaagagga 600
agattttcag tcaatgactt atggatttaa aatggctaac agtgtgacag atcttcgagt 660
tacaggcatg ctaaaagatg tggaggatga catgcaaaga agagtaaaga gtactcgaag 720
tcgacaagga gaagaaagag atccagaagt tgaactagaa caccaacaat gtttagcagt 780
attcagcaga gtgaaattta ctcgtgtgtt actgacagtg cttatagcct ttactaagaa 840
agagaccagt gctgttgcag aagctcaaaa attgatggtt caagcagcag atcttctttc 900
tgccattcat aattcattgc atcatggcat ccaggcccag aatgatacta caaaaggaga 960
tcatccaatt atgatgggtt ttgaacccct tgtgaaccag aggctacttc cacctacctt 1020
ccctcgatat gcaaaaataa ttaaaaggga agaaatggtg aactattttg caagattaat 1080
agatagaata aaaactgtct gtgaggttgt gaatttaaca aatttacatt gtatcctgga 1140

tttttttctgt gaatttagtg aacagtcacc atgtgttctt tcaagatctc tgttacaaac 1200
cactttcctg gtggataaca aaaaggtctt tggaactcat ctcatgcaag acatggtgaa 1260
agatgcactt cggtcttttg tcagtcctcc ggtgctttcc cccaagtgct acctatataa 1320
taatcaccag gctaaggact gtatcgactc ctttgttact cactgtgttc ggccattctg 1380
tagtcttatt cagatccatg gacataacag ggctcgacag agagataagc ttggtcatat 1440
tcttgaggaa tttgccacct tgcaggatga ggcagagaag gttgatgcag cgcttcacac 1500
catgctgttg aaacaggaac cccaaaggca acatttggcc tgtttaggta cctgggtcct 1560
ttaccataac cttcgcatta tgatacagta ccttctaagt ggctttgaat tggaactcta 1620
cagtatgcac gagtactatt acatatattg gtatctctct gaattccttt acgcatggtt 1680
gatgtcaaca ttgagtcgtg ccgatggctc tcaaatggca gaggaaagga taatggaaga 1740
gcagcagaaa ggccgtagta gtaaaaaaac aaagaaaaaa aagaaagttc gcccattgag 1800
ccgagagatc acaatgagcc aagcatatca gaacatgtgt gctggaatgt ttaaaaccat 1860
ggtagcattt gacatggacg gcaaagtacg taaaccgaag tttgagcttg atagtgaaca 1920
agttcggtat gaacacaggt ttgctccatt caacagtgtg atgaccccgc cgccagtgca 1980
ctacttacag ttcaaggaaa tgtctgacct caataaatat agccctcctc ctcagtctcc 2040
tgaactgtat gtggcagcta gtaagcactt tcaacaggca aaaatgatat tggaaaatat 2100
tcctaacccg gaccatgagg ttaatagaat tttaaaggtt gccaaaccca actttgtggt 2160
tatgaagtta ttggcaggag gacacaaaaa ggaatctaaa gttcctcctg aatttgattt 2220
ctctgctcat aaatattttc ctgttgtgaa acttgtttga gagagactgg ggaggtggcc 2280
ataaaggggc agagtcttct ttcagaccca actcttagag ggcacatcac caggctccac 2340
atcacgggaa gtgagatgga tttcttgggt aacaactcat tataaggaat acttttagtt 2400
tgacagcctt atatgacatg aatgaaaact gctgttttaa agtggtttat tatgttccat 2460
gtaagacact gggttccatt aatttgaa 2488

<210> 24
<211> 2647
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3072268CB1

<220>
<221> unsure
<222> 2596
<223> a, t, c, g, or other

<400> 24
cgttcgctct tcgttcggct cgagctcgag ccgaattcgg ctcgagcggc tcgagggaag 60
ccggtagccg cctcagtcgt cagaccggtg ctagcgacac gggagtgggc aaacgcatcc 120
tcttgccgtt cccggtgttt gggccttgcc tgtgacggtg ggaaaagaaa atggccttgc 180
tgtgctacaa ccggggctgc ggtcagcgct tcgatcctga gaccaattcc gacgatgctt 240
gcacatacca cccaggtgtt ccggtctttc acgatgcatt aaagggttgg tcttgctgta 300
agagaagaac aactgatttt tctgatttct taagcattgt aggctgtaca aaaggtagac 360
ataatagtga gaagccacct gagccagtca aacctgaagt caagactact gagaagaagg 420
agctatgtga attaaaaccc aaatttcagg aacacatcat tcaagcccct aagccagtag 480
aagcaataaa aagaccaagc ccagatgaac caatgacaaa tttggaatta aaaatatctg 540
cctccctaaa acaagcactt gataaactta aactgtcatc agggaatgaa gaaaataaga 600
aagaagaaga caatgatgaa attaagattg ggacctcatg taagaatgga gggtgttcaa 660
agacatacca gggtctagag agtctagaag aagtctgtgt atatcattct ggagtaccta 720
ttttccatga ggggatgaaa tactggagct gttgtagaag aaaaacttct gattttaata 780
cattcttagc ccaagagggc tgtacaaaag ggaaacacat gtggactaaa aaagatgctg 840
ggaaaaaagt tgttccatgt agacatgact ggcatcagac tggaggtgaa gttaccattt 900
cagtatatgc taaaaactca cttccagaac ttagccgagt agaagcaaat agcacattgt 960
taaatgtgca tattgtattt gaaggagaga aggaatttga tcaaaatgtg aaattatggg 1020

gtgtgattga tgtaaagcga agttatgtaa ctatgactgc aacaaagatt gaaatcacta 1080
tgagaaaagc tgaaccgatg cagtgggcaa gccttgaact gcctgcagct aaaaagcagg 1140
aaaaacaaaa agatgacaca acagattgag tgggagatgg aaggaaggct attacattat 1200
ttccgaattt ttaatactgt gtgaagtggt ggcttgctgc tgtaatcttt tgttttgttg 1260
ttgtgttact gaatgtggca tttcagggtt aacattaggt tcttaaaagc caaagtcagt 1320
ttgtcttttt gtgcctctca tctttctttt gtgttatgta agattgatta ttcatttctc 1380
cctactggta ggaaccatag ttgtgtccta tacttgaaga ggctggaaag tagcccataa 1440
ccataattgc agtatttctt tgtatttctc tgttaagcaa agaaatatta aggaacattt 1500
tttttatgtt tttgtattat tccataatta gtaaagcaag atgaaatgtc aaatttttaat 1560
cagtttttttc atggatttgt gttcttacag tacttgaaaa tatttaagga agagatgaag 1620
ctctgcagtt ttttctatgt gggatgatta ctttttttaag gaggattaat tctgaggtag 1680
tatagtaact aaaggggaat atatgaattg tttaacaaat tagaatttgt ttacaactac 1740
ttgaattttt aaattatgtc aaaacttaca ttacttgcca agcagtatga tgttatagga 1800
aacataaata agattacaga ggtatcaatt tggttaaaat tcaccatttt ataagactaa 1860
gcaataatct taacaacctc tttcctgaat atttaaatgt gtttgtatgg tgttatgact 1920
aattgttact gatttagaga ctaagccctc ttaaaacctt tagttaaata taaaaagaaa 1980
ttatatatat cttgcctccc tgatggaaaa ctatataaaa ttgtagactt aaaaggtttg 2040
tggaaataca ttaggatatc agaaaactaa atatatggag ttgctttatg actattacat 2100
gttaaataaa aatagcttaa ttgttttgga gtttttttttt aaagtcatga agagcttcac 2160
aattctagtg atacaatgtt gactatatat tgtactttat attaaatata ggtaatacag 2220
aattagatcg tttggatagg ttttgagtta gtgcagaact caagaagaaa aattaggaaa 2280
tattctgaag gccagaactc tggaagaact tagaaagttt gggaataaat taaacttatg 2340
tagatttaaa attaaaaggg gtttatttcc caaacccctt gagtcttttc ttttccttgg 2400
tatttatgga attcttacaa agaaaaattt gcagaagtac tattttcgtg aataaaacat 2460
atgtatcatt ctactaccat tagcaaataa cttagttgta gttaatataa ttgttaaaag 2520
gtttagctct gttacaaata aaaacataaa gatctggcaa ccagactttc atgatgttac 2580
attaccatga caaganactc tggggcagta aaagttggac ccccatggag tacataagca 2640
tctgata 2647

<210> 25
<211> 2337
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5519523CB1

<400> 25
aacaatttct gtatgtccag ctgagcatgc ccagtgcaat gcagttctgc tgtcctacag 60
agaagcagta atagaaacat tcttgaaata gaaggcaaaa aagcaagaaa aaaatactgg 120
tgctggcaaa tctatagttt gaaaatacga aatgcagagc actggatcct cagttttatc 180
caagtatgaa gatcagatta ctattttcac tgactaccta gaagaatatc cagatacaga 240
tgagctggta tggatcttag ggaagcagca tctccttaaa acagaaaaat ctaagctgtt 300
gtctgatata agtgctcgtc tatggtttac atacagaagg aaattttcac caattggtgg 360
aacgggccct tcatcagatg ctggttgggg atgtatgcta cgctgtggac agatgatgct 420
ggctcaagcc cttatctgta gacacttggg aagggactgg agctgggaga aacaaaaaga 480
acaacccaaa gaataccaac gcatcctaca gtgcttctta gatagaaaag attgttgcta 540
ctctatccat caaatggcac aaatgggtgt aggagaaggg aaatcaattg gagaatggtt 600
tggaccaaat acagttgcac aggtgttaaa aaaacttgct ttatttgacg aatggaattc 660
cttggctgtt tatgtttcaa tggataacac agtggtcatt gaagatatca aaaaaatgtg 720
ccgtgtcctt cccttgagtg ctgacacagc tggtgacagg cctcccgatt ctttaactgc 780
ttcaaaccag agtaagggca cctctgccta ctgctcagcc tggaaacccc tgctgctcat 840
tgtgcccctt cgcctgggca taaaccaaat caatcctgtc tatgttgatg cattcaaaga 900
gtgttttaag atgccacagt ctttaggggc attaggagga aaaccaaata acgcgtatta 960
tttcatagga ttcttaggtg acgagctcat cttcttggac cctcatacaa cccagacctt 1020

tgttgacact gaagagaatg gaacggttaa tgaccagact ttccattgcc tgcagtcccc 1080
acagcgaatg aacatcctaa acctggatcc ttcagttgca ttgggatttt tctgcaaaga 1140
agaaaaagac tttgataact ggtgtagcct tgttcagaag gaaattctaa aggagaattt 1200
aaggatgttt gaattagttc agaaacatcc atcacactgg cctccctttg tacctccagc 1260
caagccagaa gtgacaacca ctggggcaga attcattgac tctactgagc aactggagga 1320
gtttgatctg gaggaagatt ttgagattct gagtgtgtag aatcctggga actcaacttg 1380
aaggtctgtc ttccatctgg caccataaaa acatgaactt attgcataaa acttttctag 1440
tcagcaagtg cctgatatgc caatagcata caaactcaat agcaatcatg actgagccaa 1500
tcactgtttc tcagaaaaac aaaacaaaac aaaacaaatg acagtaaccc ttccccggaa 1560
agaaatagaa caatcatgga gcctaggagc agagagatga ggaggagttc attgcttccc 1620
agcttgtgtt atatggctac agcaagtctt cagctgctgc aatgaggaaa tgggcatctg 1680
gaagacaaac agcaactctc agcttgcttc aagaaccagc agataagaga tggttaagct 1740
gttcttcacc ctttcagatg tgacctcttt tggactaagc agcaatctgt tctcttgctc 1800
aaataataaa gtgactgaat cagggaggaa aaggttcttg ttaaattatt tgattgtgta 1860
gttgaagtaa ttataattta tatcaaaacg tttgtcaaag aaacgatgtc aaatatacac 1920
ttcttgatct cccttctgtt tgcggggatc ttactatttg atgggtcact gtccccattc 1980
ttactgatac ttttgtcaga tatcaccctg tccttaaatc atgatcactt aaatcagggg 2040
tcagcaaact ttttctgtaa agggccagac gggaaatatt ttgggctttg caggccatgc 2100
ggcctctgtc acatctactc aactctgctg ttgacatgca aaagcagcaa tagacaatat 2160
gcgtgtaaat gagtgtggct gtaatccaag aaaactttat ttacaaaagc aggtggaggg 2220
ctgggtttgg cctgcaggct gtagcttgcc aatcagtgac ttaaattgtt gatttttgtt 2280
tgataaatta aaaataaatt gtgtttgtgt atataggatc ctgtagattt acgtatg 2337

<210> 26
<211> 3141
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1760208CB1

<400> 26
ggagggcaga ggccagggtc tgggatgcct tggagggtaa gtgggaggga tgatgatcct 60
ggctggaagt ccacctgctg ctgcctgcct gccacttaga gccttctcat tcctattcca 120
ggctgtctgt ccccagaccc cagagcacgt ccggcaccac catgactggg ctgttgaaga 180
ggaaatttga ccagctggat gaggacaact cctcggtctc ctcctcctcc tcttcctctg 240
ggtgccagtc tcgctcctgc tccccaagct cttctgtctc ccgtgcctgg gactcagagg 300
aggaaggccc ctgggatcag atgcccctgc ctgaccgtga cttctgcggc cccagaagtt 360
tcaccccccc gtctatcctg aagcgagctc gccgggagcg cccaggccgt gtagcctttg 420
atgggatcac cgtcttctac ttcccccgct gccagggctt caccagtgtg cccagccgtg 480
gtggctgtac tctgggtatg gcccttcgcc acagtgcttg ccgtcgcttc tctttggctg 540
agtttgcgca ggagcaagcc cgtgcacggc acgagaagct ccgccagcgc ttgaaagagg 600
agaagttgga gatgctgcag tggaagcttt cggcagctgg ggtaccccag gcagaggcag 660
ggctgccacc tgtggtggat gccattgatg acgcctctgt ggaggaggac ttggcagtcg 720
ctgtggcagg tggccggttg gaagaagtga gcttcctaca gccctaccca gcccggcgac 780
gtcgagctct gctgagggct tcaggtgtgc gaaggatcga tcgggaggag aagcgggagc 840
tgcaggcact gcgccaatcc cgggaggatt gtggctgtca ctgcgatagg atctgcgacc 900
ctgagacctg cagctgcagc ctggcaggca tcaagtgcca gatggaccac acagcattcc 960
cctgtggctg ctgcagggag ggctgtgaga accccatggg ccgtgtggaa tttaatcagg 1020
caagagttca gacccatttc atccacacac tcacccgcct gcagttggaa caggaggctg 1080
agagctttag ggagctggag gcccctgccc agggcagccc acccagccct ggtgaggagg 1140
ccctggtccc tactttccca ctggccaagc cccccatgaa caatgagctg ggagacaaca 1200
gctgcagcag cgacatgact gattcttcca cagcatcttc atcagcatcg ggcactagtg 1260
aggctcctga ctgccccacc cacccaggcc tgcctggccc tggcttccag cctggcgttg 1320
atgatgacag cctggcacgc atcttgagtt tcagtgactc tgacttcggt ggggaggagg 1380

aggaagagga ggaagggagt gtggggaacc tggacaacct cagctgcttc catccagctg 1440
acatctttgg tactagtgac cctggtggcc tggccagctg gacccacagc tattctggct 1500
gtagcttcac atcaggcatc ctggatgaga atgccaacct ggatgccagc tgcttcctaa 1560
atggtggcct tgaagggtca agggaaggca gccttcctgg cacctcagtg ccacccagca 1620
tggacgctgg ccggagtagc tcagtggatc tcagcttgtc ttcttgtgac tcctttgagt 1680
tactccaggc tctgccagat tatagtctgg ggcctcacta cacatcacag aaggtgtctg 1740
acagcctgga caacatcgag gcacctcact tccccctgcc tggcctgtct ccacctgggg 1800
atgccagcag ttgcttcctg gagtccctca tgggcttctc cgagccagcc gccgaagccc 1860
tagatccctt tattgacagc cagtttgagg acactgtccc agcatctcta atggagcctg 1920
tgccggtgtg aggaccagga tgtcttttcc cagccccaag agacctgttg ctgctttctt 1980
gtaattatgg ggctccccag agtctgcgta acagtctccc actggctggc tcacccacag 2040
gtgccatgtg cacactcctg gttttcaaac aattctctgg atttatttat ttgttttaac 2100
ttttctgtgc tgaagagagg actaggggga gggggcttcc cctttcagct gcccggcccc 2160
ccacacccac agcttgctct tctatctcca caacgtgagc ctggaagagg agaaaatgtg 2220
gctcctctgg agcttggcag accacttttc ggtctttgcg tgatgttcct tagcccaaag 2280
acggtgagac agggctgaaa tcaggtggct tctgccaccc tgagccctag acccatgggt 2340
ggctaaatcc actggactgt gaagactata atttatttcc ataatttatt tggagattga 2400
ggaggctttg gttgcacttc tttggctggt gggtaatgcc aggggtgggg tgggcacagg 2460
ccctcaagag ccccttttgc cttgtagtcc tacaccttgc cctgcctggg ctttggtgca 2520
gactaggtgt ggatttgagc tctgtgatct atgtctgctg cctggctcct agatggctct 2580
gcgggcaggt gctggccaag gacatcatct aggcaggggg agagcctggg ctgaacagct 2640
gtgaccaaaa ctcccttctg ccccaccctg ccccctccac ttcctgccct ctgttccatc 2700
ttccccttc ccaaaggcca cagcctttat tccaggccca gggatgtagg agggggaagg 2760
aggaaacagg aagcccagag agggcaaagg gcctacctcg gggcgcgaac catgccccag 2820
actattatct cagggctttc tgggcactgc acttcagcgt ggcccacctg cccatgccct 2880
gaggccagtt ggcgaggggt ggctcctgag ggtttttata ccctttgttt gctaatgttt 2940
aattttgcat cataatttct acattgtccc tgagtgtcag aactataatt tattccattt 3000
ctctctgtgt ctgtgccaag aaacgcaggc tctgggcctg ccccttgccc aggaggcctt 3060
gccagcctgt gtgcttgtgg gaacaccttg tacctgagct tacaggtacc aataaagagg 3120
ctttatttt aaaaaaaaaa a 3141

<210> 27
<211> 3261
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1900132CB1

<400> 27
gatccctgag ctcgtcaagc agctgtcggc caagctcatc gccaacgatg acatggcaga 60
gcttctcctc ggggagtcga agctggagca gtacctgaag gaacaccccc tgaggcaggg 120
ggccacgtcc ccgaggcccc aagccccagc tgactgaggt ccgcaagcat ctgaccgccg 180
ccctggaccg agggaacctt aagtcagaat tcctacaaga atccaatctg atcatggcca 240
agttgaatta tgtggaaggt gattataaag aagctctgaa catttacgcc cgggtgggcc 300
tggacgatct gccactgaca gctgtcccgc cctacaggct gcgggtgatc gcagaagcct 360
acgctaccaa aggactttgt ttggagaagc tgcctatttc ttcttctacc agtaatctcc 420
atgtggaccg ggaacaggat gtcatcacct gttatgagaa agcaggggac atcgcactcc 480
tgtatctcca agagatagaa agggtaatac tttctaatat tcaaaacaga agccctaagc 540
ctggccctgc tccccacgat caagaactag gttttttcct agaaacagga cttcagagag 600
cccatgtcct ctatttcaaa aatgggaact tgacaagagg agtcggaaga tttagagagc 660
ttctcagagc agttgaaaca agaacaactc aaaacctgcg aatgacaata gccaggcagc 720
tggcagagat cttgttgcgg ggtatgtgtg agcagagcta ctggaaccct ctggaggatc 780
caccgtgcca gtcacctctg gacgatcctc tccgcaaagg agcaaacaca aaaacctaca 840
ctctcactcg gagagcccgt gtctactcag gagagaacat tttttgtcct caagaaaata 900

cggaagaagc cctgttgtta ttgctgatta gtgaatcaat ggccaaccgg gacgctgtgc 960
tgagcaggat acctgaacac aagagtgacc gcctcatcag tctgcagagt gcatctgtgg 1020
tctatgactt actcaccatt gctcttggaa gaagaggcca gtatgagatg ctgtcagagt 1080
gcctagaaag agccatgaag tttgcctttg aggaattcca cctgtggtac cagtttgctc 1140
tgtccctgat ggctgctgga aaatctgccc gtgccgtgaa ggtgctgaaa gagtgtatcc 1200
gcctgaagcc ggacgatgcc accatccctc tcctcgctgc caagctctgc atgggctccc 1260
tgcactggtt ggaagaggct gaaaagtttg ccaaaactgt cgttgatgtg ggagagaaaa 1320
cgtcagagtt caaggccaaa ggctacttag ctctggggct cacgtacagt ctgcaggcca 1380
ctgacgcttc tttgcgaggg atgcaggagg tcctacagag aaaggcgctt cttgcatttc 1440
agagggccca cagcctgtca cccacagatc accaagcagc tttctacctg gctctgcagc 1500
ttgccatctc cagacagatc ccagaggctc tggggtatgt ccgccaagct cttcagcttc 1560
aaggtgacga tgccaactcc ctgcacctcc ttgccctcct gctgtcagca cagaagcatt 1620
accatgacgc tctgaacatc atcgacatgg ccctgagtga atacccagaa aatttcatac 1680
tactgttttc caaagtgaag ttgcagtcac tctgccgagg cccggacgag gcactgctga 1740
cttgtaagca catgctgcag atatggaaat cctgctacaa cctcaccaac cccagtgatt 1800
ctggacgtgg gagcagcctc ttagatagaa ccattgctga cagacgacag cttaatacaa 1860
ttactttgcc agacttcagc gatcccgaga caggctccgt ccatgccaca tcggtagcag 1920
cctcaagagt ggagcaggca ctgtcggaag tggcttcgtc tctgcagagc agtgccccta 1980
agcagggccc gctgcacccc tggatgacgc tggcacagat ctggctccat gcagctgaag 2040
tctatatcgg catcgggaag cctgcagaag ccacagcctg tacccaagaa gctgccaacc 2100
tcttcccaat gtcccacaat gtcctctaca tgcgcggcca gattgctgag ctccggggaa 2160
gcatggacga ggcgcggcgg tggtatgaag aggccttagc catcagcccc acccacgtga 2220
agagcatgca gcgactggcc ctgatccttc accagctagg ccgctacagt ctggcggaga 2280
agatcctccg ggacgcggtg caggtgaact cgacagccca cgaggtctgg aacgggctgg 2340
gcgaggtcct ccaagctcag ggcaacgatg cggcggctac ggagtgcttc ctgacagcct 2400
tggagctgga ggccagcagc cccgccgtgc ccttcaccat catcccccgc gtgctctgag 2460
caggcgcctg ccagcctcac ctgccgctca ggcctcagag gccctgccgg gcaccagggc 2520
ttgtgccatc gccccaaggg gatgaatctg ccgcactgag gccagggacg agtgttcagt 2580
gggccacagt gaaccaacca aaccaacccc gaatcatcgc tctcgccatg tgcgtttctc 2640
ttgttttttt tgccagccca atggtagttt ctgaacctat tgacattgtt caaaatggat 2700
catgtgccat attttgttag ttgacatctg agttttcagt aaaatgatta tggaattaat 2760
cagcaaatgt agaagaatat attcaaagtt aaaattcagt ggcagcacag attatttta 2820
tcagagctgt aaagaaaaca actgtccttt tctccccacc acccctcctg ccccactttg 2880
gcccagaaac caaatgtgaa cttcctgtct cccacctcag cactagtcca tgccaggaca 2940
ccagctgaca atttcttggt tttactgtca ataattgtac catgtgatca attactgtcc 3000
tcacttagaa caaagcctga gtccgagaat atttatattt taccaatata tgcctgttac 3060
aagagaagga aatatgagtt atttaagttt aacttttta tgtgaattca gagtttattt 3120
atcgagggaa atatgtacaa agaagcttca aatggaatat ttaccgacat tccttataca 3180
tgacagacac ttggctacat gggaagatga tgttaataat aaaatgattt ttaaatgaaa 3240
aaaaaaaaaa aaaaaactcg g 3261

<210> 28
<211> 1097
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7487551CB1

<400> 28
cgatcactga cgagctcgca tcacttataa cggcgcagtg tgctggaaag tgattgggtg 60
tggaggagtg agagggaggc gggcgtcagg ggtagctcca aggtttaact taggtgactt 120
cagatctcca atcaccaagc cctctctggt cctgccttct ccacctgctc ctgcgggtct 180
tgcatcttct cctgtgtacc tccagtgagg agtggtcccc accaccctcc ccatcagtgc 240
acttacgaag tgctctcatc ttcacaaaca agccagcacc cagcccagcc ctggtagtca 300

gggcggttgc cacagcaatt gacatcagcg acctggtccc caaggaacct gccaccttcc 360
gcctgcctgc agggcctgca ttatcgcttc tgcggggact ggagtggagg cagatgggga 420
ctcccacccc tgacacacac cccattttga gaactgagtg gggctgggaa gagccagtgg 480
caaagggagg ggaagaggga agggcagaaa gtaggtgggg ccccccttg gtggcctctt 540
ctctccacgg ccccaggctc cagcccactt gggtccttgg cgttggtggc agcagcactt 600
gggccatggc ggaggacagg ccgcagcagc cgcagctgga catgccgctg gtcctggacc 660
agggcctgac caggcagatg cggctacgcg tggagagcct gaagcagcgc ggggagaagc 720
gccaggatgg ggagaagctg ctgcagccag cggagtctgt gtaccgcctc aacttcaccc 780
agcagcagcg gctacagttc gagcgctgga atgtcgtgct ggacaagccg ggcaaggtca 840
ccatcacagg cacctcgcag aactggacgc ctgacctcac caacctcatg acacgccagc 900
tgctggaccc cactgccatc ttctggcgca aggaggactc ggatgccata gattggaatg 960
aggccgacgc cctggagttt ggggagcgcc tgtcggacct ggccaagatc cgcaaggtca 1020
tgtacttcct cgtcaccttt ggcgagggtg tggagcccgc caacctcaag gcctccgtgg 1080
tttttaacca gctctga 1097

<210> 29
<211> 1633
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1871014CB1

<400> 29
ctaccccaca atccttagct ctttccgtct ccactcggct tccgtccatt cttccggtgg 60
agatggctgc ggccgtggcg gggatgctgc gaggggtct cctgccccag gcgggccggc 120
tgcctaccct ccagactgtc cgctatggct ccaaggctgt tacccgccac cgtcgtgtga 180
tgcactttca gcggcagaag ctgatggctg tgactgaata tatccccccg aaaccagcca 240
tccacccatc atgcctgcca tctcctccca gccccccaca ggaggagata ggcctcatca 300
ggcttctccg ccgggagata gcagcagttt tccaggacaa ccgaatgata gccgtctgcc 360
agaatgtggc tctgagtgca gaggacaagc ttcttatgcg acaccagctg cggaaacaca 420
agatcctgat gaaggtcttc cccaaccagg tcctgaagcc cttcctggag gattccaagt 480
accaaaatct gctgcccctt tttgtggggc acaacatgct gctggtcagt gaagagccca 540
aggtcaagga gatggtacgg atcttaagga ctgtgccatt cctgccgctg ctaggtggct 600
gcattgatga caccatcctc agcaggcagg gctttatcaa ctactccaag ctccccagcc 660
tgcccctggt gcaggggggag cttgtaggag gcctcacctg cctcacagcc cagacccact 720
ccctgctcca gcaccagccc ctccagctga ccaccctgtt ggaccagtac atcagagagc 780
aacgcgagaa ggattctgtc atgtcggcca atgggaagcc agatcctgac actgttccgg 840
actcgtagcc agcctgttta gccagccctg cgcataaata cactctgcgt tattggctgt 900
gctctcctca atgggacatg tggaagaact tggggtcggg gagtgtgttt gtcacttggt 960
tttcactagt aatgatattg tcaggtatag ggccacttgg agatgcagag gattccattt 1020
cagatgtcag tcaccggctt cgtccttagt tttcccaact tgggacgtga taggagcaaa 1080
gtctctccat tctccaggtc caaggcagag atcctgaaaa gatagggcta ttgtcccctg 1140
cctccttggt cactgcctct tgctgcacgg gctcctgagc ccacccccct ggggcacaac 1200
ctgccactgc cacagtagct caaccaagca gttgtgctga gaatggcacc tggtgagagc 1260
ctgctgtgtg ccaggctttg tgctgagtgc tgtacgtgta ttagttcctt tactgctgac 1320
cacattgtac ccatttcaca gagaaggagc agagaaatta agtggcttgc tcaaggtcat 1380
gcagttagta agtggcagaa cagggacttg aaccaagccc tctgctctga agaccgcgtc 1440
ctgaatttct acactagagc ttcctcatca ggttacccag aagtgggtcc catccaccat 1500
ccaggtgtgc ttggatgttc tccaccctcg aggtgtacgc tgtgaaaagt ttgggagcac 1560
tgctttataa taaaatgaaa tatactactt cctttaaaaa aaaaaaaaaa aaaaaaaaaa 1620
aaaaaaaaaa aaa 1633

<210> 30
<211> 5869

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2903166CB1

<400> 30
cgtgttaata tagagcagcg tgagtcgcgt tgcatgacgt acgtacgcgc gggaagtacg 60
gctcgaggtt agacaaggaa tggtcatata tctttaaagt tccctaagaa aagctaggca 120
cagagaggtt aagtgacttg cccaagatcc tacctgcagc aactgtcaga tccaggcttg 180
aacttaacat ttgttttcta taactgacta taactccata actccctgcc ctctgcagat 240
gctgagcaag ggtctgaagc ggaaacggga ggaggaggag gagaaggaac ctctggcagt 300
cgactcctgg tggctagatc ctggccacac agcggtggca caggcacccc cggccgtggc 360
ctctagctcc ctctttgacc tctcagtgct caagctccac cacagcctgc agcagagtga 420
gccggacctg cggcacctgg tgctggtcgt gaacactctg cggcgcatcc aggcgtccat 480
ggcacccgcg gctgccctgc cacctgtgcc tagcccacct gcagccccca gtgtggctga 540
caacttactg gcaagctcgg acgctgccct ttcagcctcc atggccagcc tcctggagga 600
cctcagccac attgagggcc tgagtcaggc tccccaaccc ttggcagacg aggggccacc 660
aggccgtagc atcgggggag cagcgcccag cctgggtgcc ttggacctgc tgggcccagc 720
cactggctgt ctactggacg atgggcttga gggcctgttt gaggatattg acacctctat 780
gtatgacaat gaactttggg caccagcctc tgagggcctc aaaccaggcc ctgaggatgg 840
gccgggcaag gaggaagctc cggagctgga cgaggccgaa ttggactacc tcatggatgt 900
gctggctccg ggcatgaccc tcacagccac gggcctggga cagagagctg atgacccagg 960
agacccccctc tactaccacc tacaaggttc aggcttctcg tgtccccagc tcaggactct 1020
gtgctgtgta tcagtcctgg agcgccggac ccaggaggcc caaggagctg gaggtgaccc 1080
tcaggcagca agaaccccca cggaagggcg tgagccctgc agacagctgt gcggcacctc 1140
gggctgggct cctgttagga ggaagtgcct gcacccaggc agcggctcag aggcagctgc 1200
tccatgcaga actgaagctg gttctgcagc agaaagggga gaggacacag gagcctgggg 1260
tgcaggtgcc tcccagcaac gccatggagg ccaggagccg gagtgccgag gagctgaggc 1320
gggcggagtt ggtggaaatt atcgtggaga cggaggcgca gaccggggtc agcggcatca 1380
acgtagcggg cggcggcaaa gagggaatct tcgttcggga gctgcgcgag gactcacccg 1440
ccgccaggag cctcagcctg caggaagggg accagctgct gagtgcccga gtgttcttcg 1500
agaacttcaa gtacgaggac gcactacgcc tgctgcaatg cgccgagcct tacaaagtct 1560
ccttctgcct gaagcgcact gtgcccaccg gggacctggc tctgcggccc gggaccgtgt 1620
ctggctacga gatcaagggc ccgcgggcca aggtggccaa gctgaacatc cagagtctgt 1680
cccctgtgaa gaagaagaag atggtgcctg gggctctggg ggtccccgct gacctggccc 1740
ctgttgacgt cgagttctcc tttcccaagt tctcccgcct gcgtcggggc ctcaaagccg 1800
aggctgtcaa gggtcctgtc ccggctgccc ctgcccgccg gcgcctccag ctgcctcggc 1860
tgcgtgtacg agaagtggcc gaagaggctc aggcagcccg gctggccgcc gccgctcctc 1920
cccccaggaa agccaaggtg gaggctgagg tggctgcagg agctcgtttc acagcccctc 1980
aggtggagct ggttgggccg cggctgccag gggcggaggt gggtgtcccc caggtctcag 2040
cccccaaggc tgccccctca gcagaggcag ctggtggctt tgccctccac ctgccaaccc 2100
ttgggctcgg agccccggct ccgcctgctg tggaggcccc agccgtggga atccaggtcc 2160
cccaggtgga gctgcctgcc ttgccctcac tgcccactct gcccacactt ccctgcctag 2220
agacccggga aggggctgtg tcggtagtgg tgcccaccct ggatgtggca gcaccgactg 2280
tgggggtgga cctggccttg ccgggtgcag aggtggaggc ccggggagag gcacctgagg 2340
tggccctgaa gatgccccgc cttagttttc cccgatttgg ggctcgagca aaggaagttg 2400
ctgaggccaa ggtagccaag gtcagccctg aggccagggt gaaaggtccc agacttcgaa 2460
tgcccacctt tgggctttcc ctcttggagc cccggcccgc tgctcctgaa gttgtagaga 2520
gcaagctgaa gctgcccacc atcaagatgc cctcccttgg catcggagtg tcagggcccg 2580
aggtcaaggt gcccaaggga cctgaagtga agctccccaa ggctcctgag gtcaagcttc 2640
caaaagtgcc cgaggcagcc cttccagagg ttcgactccc agaggtggag ctccccaagg 2700
tgtcagagat gaaactccca aaggtgccag agatggctgt gccggaggtg cggcttccag 2760
aggtagagct gcccaaagtg tcagagatga aactcccaaa ggtgccagag atggctgtgc 2820
cggaggtgcg gcttccagag gtacagctgc tgaaagtgtc ggagatgaaa ctcccaaagg 2880

tgccagagat ggctgtgccg gaggtgcggc ttccagaggt acagctgccg aaagtgtcag 2940
agatgaaact cccagaggtg tcagaggtgg ctgtgccaga ggtgcggctt ccagaggtgc 3000
agctgccgaa agtgccagag atgaaagtcc ctgagatgaa gcttccaaag gtgcctgaga 3060
tgaaacttcc tgagatgaaa ctccctgaag tgcaactccc gaaggtgccc gagatggccg 3120
tgcccgatgt gcacctccca gaagtgcagc ttccaaaagt cccagagatg aagctccctg 3180
agatgaaact ccctgaggtg aaactcccga aggtgcccga gatggctgtg cccgatgtgc 3240
acctcccgga agtgcagctc ccgaaagtcc cagagatgaa actccctaaa atgcctgaga 3300
tggctgtgcc agaggttcga ctccccgagg tgcagctgcc aaaagtctca gagatgaaac 3360
tccccaaggt gcctgaaatg gccgtgcccg atgtgcacct cccagaggtg cagctgccca 3420
aagtctgtga aatgaaagtc cctgacatga agctcccaga gataaaactc cccaaggtgc 3480
ctgagatggc tgtgccccgat gtgcacctcc ccgaggtgca gctgccgaaa gtgtcagaga 3540
ttcggctgcc ggaaatgcaa gtgccgaagg ttcccgacgt gcatcttccg aaggcaccag 3600
aggtgaagct gcccagggct ccggaggtgc agctaaaggc caccaaggca gaacaggcag 3660
aagggatgga atttggcttc aagatgccca agatgaccat gcccaagcta gggagggcag 3720
agtccccatc acgtggcaag ccaggcgagg cgggtgctga ggtctcaggg aagctggtaa 3780
cacttccctg tctgcagcca gaggtggatg gtgaggctca tgtgggtgtc ccctctctca 3840
ctctgccttc agtggagcta gacctgccag gagcacttgg cctgcagggg caggtcccag 3900
ccgctaaaat gggcaaggga gagcgggcgg agggccccga ggtggcagca ggggtcaggg 3960
aagtgggctt ccgagtgccc tctgttgaaa ttgtcacccc acagctgccc gccgtggaaa 4020
ttgaggaagg gcggctggag atgatagaga caaaagtcaa gccctcttcc aagttctcct 4080
tacctaagtt tggactctcg gggccaaagg tggctaaggc agaggctgag ggggctgggc 4140
gagctaccaa gctgaaggta tccaaatttg ccatctcact ccccaaggct cgggtggggg 4200
ctgaggctga ggccaaaggg gctggggagg caggcctgct gcctgccctc gatctgtcca 4260
tcccacagct cagcctggat gcccacctgc cctcaggcaa ggtagaggtg gcaggggccg 4320
acctcaagtt caaggggccc aggtttgctc tccccaagtt tggggtcaga ggccgggaca 4380
ctgaggcagc agaactagtg ccaggggtgg ctgagttgga gggcaagggc tggggctggg 4440
atgggagggt gaagatgccc aagctgaaga tgccttcctt tgggctggct cgagggaagg 4500
aagcagaagt tcaaggtgat cgtgccagcc cgggggaaaa ggctgagtcc accgctgtgc 4560
agcttaagat ccccgaggtg gagctggtca cgctgggcgc ccaggaggaa gggagggcag 4620
agggggctgt ggccgtcagt ggaatgcagc tgtcaggcct gaaggtgtcc acagccaggc 4680
aggtggtcac tgagggccat gacgcggggc tgaggatgcc tccgctgggc atctccctgc 4740
cacaggtgga gctgaccggc tttggggagg caggtacccc agggcagcag gctcagagta 4800
cagtcccttc agcagagggc acagcaggct acagggttca ggtgccccag gtgaccctgt 4860
ctctgcctgg agcccaggtt gcaggtggtg agctgctggt gggtgagggt gtctttaaga 4920
tgcccaccgt gacagtgccc cagcttgagc tggacgtggg gctaagccga gaggcacagg 4980
cgggcgaggc ggccacaggc gagggtgggc tgaggctgaa gttgcccaca ctgggggcca 5040
gagctagggt ggggggcgag ggtgctgagg agcagccccc aggggccgag cgtaccttct 5100
gcctctcact gcccgacgtg gagctctcgc catccggggg caaccatgcc gagtaccagg 5160
tggcagaggg ggagggagag gccggacaca agctcaaggt acggctgccc cggtttggcc 5220
tggtgcgggc caaggagggg gccgaggagg gtgagaaggc caagagcccc aaactcaggc 5280
tgccccgagt gggcttcagc caaagtgaga tggtcactgg ggaagggtcc cccagccccg 5340
aggaggagga ggaggaggag gaagagggca gtggggaagg ggcctcgggt cgccggggcc 5400
gggtccgggt ccgcttgcca cgtgtaggcc tggcggcccc ttctaaagcc tctcgggggc 5460
aggagggcga tgcagccccc aagtcccccg tcagagagaa gtcacccaag ttccgcttcc 5520
ccagggtgtc cctaagcccc aaggcccgga gtgggagtgg ggaccaggaa gagggtggat 5580
tgcgggtgcg gctgcccagc gtggggtttt cagagacagg ggctccaggc ccggccagga 5640
tggagggggc tcaggctgcg gctgtctgaa gcccctagtc agatggggat cccttcttgc 5700
cttcctttct ctacccccctc gctgttgtgt gtgtgataac tagcactaac cctaagaggg 5760
ccgggaggtg ggtgactgac cagggctggc agggaggcct gctcctgtct ctctggcagg 5820
agtgcctgta ccccaccaag ccatgtgaat aaaataatct ggaagcaaa 5869

<210> 31
<211> 3879
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1723804CB1

<400> 31
tggtaagaaa gcagagatcg aggtgacagg cgagctggct ggactcggag cgcggtcgag 60
gctttctgcg ttcgcggcgg cggaatggcc cgtgcgcggc tcgccgcgtc gcggctctgt 120
ggtccctaga cgtcggctcc cgccctcggc gctgatctcc ggcgcgggca ctgctttcca 180
ctcggctcct gtcgtccgtt ctctcaggct cccgttcagg atttttagac tctgaggagc 240
agttggagct aatccacatt atggaaatgg aaaccaccga acctgagcca gactgtgtag 300
tgcagcctcc ctctcctcct gatgactttt catgccaaat gagactctct gagaagatca 360
ctccattgaa gacttgtttt aagaaaaagg atcagaaaag attgggaact ggaaccctga 420
ggtctttgag gccaatatta aacactcttc tagaatctgg ctcacttgat ggggttttta 480
gatctaggaa ccagagtaca gatgagaaca gcttacatga acctatgatg aagaaagcca 540
tggaaatcaa ttcatcatgc ccaccagcag aaaataatat gtctgttctg attcctgata 600
ggacaaatgt tggggaccag ataccggaag cccatccttc cactgaagct ccagaacgag 660
tggttccaat ccaagatcac agctttccat cagaaaccct cagtgggacg gtggcagatt 720
ccacaccagc tcacttccag actgatcttt tgcacccagt ttcaagtgat gttcctacta 780
gtcctgactg cttagacaaa gtcatagatt atgttccagg cattttccaa gaaaacagtt 840
ttacaatcca atacattctg gacaccagtg ataagctgag tactgagctc tttcaggaca 900
aaagtgaaga ggcttccctt gacctcgtgt ttgagctggt gaaccagttg cagtaccaca 960
ctcaccaaga gaacggaatt gaaatttgca tggactttct gcaaggcact tgtatttatg 1020
gcagggattg tttgaagcac cacactgtct tgccatatca ttggcagatc aaaaggacaa 1080
ctactcaaaa gtggcagagt gtattcaatg attctcagga gcacttggaa agattttact 1140
gtaacccaga aaatgataga atgagaatga agtatggagg acaagaattt tgggcagatt 1200
tgaatgccat gaacgtgtat gaaacaactg aatttgacca actacgaagg ctgtccacac 1260
caccctctag caatgtcaac tctatttacc acacagtctg gaaattcttc tgtagggacc 1320
actttggatg gagagagtat cccgagtctg tcattcgatt gattgaagaa gccaactctc 1380
ggggtctgaa agaggttcga tttatgatgt ggaataacca ctacatcctc cacaattcat 1440
tcttcaggag agagataaaa aggagacccc tcttccgctc ctgttttata ctgcttccat 1500
atttacagac acttggtggg gttcccacac aagctcctcc acctcttgaa gcaacttcat 1560
catcacaaat tatctgccca gatggggtca cttcagcaaa cttttaccct gaaacttggg 1620
tttatatgca tccatctcag gacttcatcc aagtccctgt ttctgcagag gataaaagtt 1680
atcggatcat ttacaatctt tttcataaga ctgtgcctga gtttaaatac agaattttgc 1740
agatattgag agtccaaaac cagtttcttt gggagaaata taaaaggaaa aaggaatata 1800
tgaacaggaa aatgtttggc cgtgacagga taataaatga gagacattta tttcatggaa 1860
catcccagga tgtggtagat ggaatctgca aacacaactt tgaccctcga gtctgtggaa 1920
agcatgctac aatgtttgga caaggcagtt attttgcaaa gaaggcaagc tactctcata 1980
acttttctaa gaagtcctcc aaaggagtcc acttcatgtt tctggccaaa gtgctgacgg 2040
gcagatacac aatgggcagt catggcatga gaaggccccc gccagtcaat cctggcagtg 2100
tcaccagtga cctttatgac tcttgtgtgg ataatttctt tgagcctcag atttttgtca 2160
tttttaatga tgaccagagt tacccttatt ttgttatcca atatgaagaa gtcagtaaca 2220
ctgtttccat ttgaaaaatc ttggtactgc taaattattt gatatgaact caatccagca 2280
tttgtagcag gttttgaatg ggtgggactg ggtggggaac agcattggac attaataggg 2340
cacttttcag acccatttttt taaagtgcta gaaaatgctt ttttttaaaa aaaaatacaa 2400
gttttaaaat gaccacttac tctttaatta tttactaatt gctagtgtac tcagtgtgga 2460
aaagactaca gattacacac tcttttcatt cacacttgta catatagaca gcaatgttat 2520
taggagcatt aaattaaaaa actgaacagc ctaatttaaa tgtggcttgg gcctggtaga 2580
agtttgacca aatggaatgg aggctgtgag caatgtgagg attctattta tttatttacg 2640
tttgataaaa cttactggaa ctagtactac catgcgtatt ccctgtccaa agcatcactg 2700
ctttggtata gtataagttc atgaaattct ggtgggtaga aagaaatttt tatttctatc 2760
agcagtacta aaatgtatca gccaaccaga gaacatcagt gactttaact tctgcagagt 2820
ttgccccaga attcagagtt ctatttagag gaagttaaaa caacaacaaa aaacaaccat 2880
ttgaaaaatt tttgtcacca gcaaaacttt tcactaatta gtgatatgaa atgtgatttt 2940
tgtgttgtta aacttcagct ttggaaaact cagtctcttt cattatcatc cattccaatt 3000
tgaaggagtt gggcagctaa tttggttaaa ggcagtcttg agggttagaa gtattacttc 3060

cttttcgggt tccagaccta gcttgtgact gaaagtttta gaaaaaggaa gtacatctat 3120
agccgaattg ataaggttat tactgtgttt tgcacaaagt atattagcaa aagtatttgc 3180
tggaattatt ggtagatggc agtcccactc ctacacctgc tttgtcagat acagctgggt 3240
tccctggctg actctgtacc ttacttacac tacttactta atagaaacac aaacttggaa 3300
attgtgccag tggtccagct ggagcacaac gtttggtgaa tatgctgttt cctcagttca 3360
gagaggtagc aactaggtaa actccttata aaaagcaaat acctggattt acaaaagtga 3420
aagtagttgt tcacaaaaga attcgccatg gaattctttc agttaccaag ctctcctggt 3480
aatgtttgtg gttatatcat ttacacaaaa cttttcagga acttctgtgt tgtttaagca 3540
agatgtatct gtactgatgt ctcagtgaat cagtctgttt attaagcact tatcagggct 3600
tccacacact tatttatttt gccctagtta atcctgttgt ttgctgccat tggcatgaaa 3660
tggccaactg tggctgttac agttctttca ttcaattata acttgtaaac cagtgactcc 3720
taatcttttt caagttaaga caccttacca ttgcttattt ggttttatga gacttgttcc 3780
ttttttttctc cctaaggaaa aagaaagctt tatgacatat ttatttttt aataaaacta 3840
agcaaaaata aaacttatgg taattcttta aaaaaaaaa 3879

<210> 32
<211> 2160
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7736769CB1

<400> 32
atggccgcgg cggtggcggt ggcggccgcg tcccggcggc agtcgtgcta cctgtgtgac 60
ctgccccgca tgccctgggc catgatctgg gacttcaccg aacccgtctg ccgcggctgc 120
gtcaactacg agggcgccga ccgcgtcgag ttcgtcatcg agacggcgcg gcagctcaag 180
cgggcgcacg gctgcttccc ggagggtcgc tccccacccg gcgccgcggc ctcggccgcc 240
gccaagccgc cgccgctctc cgccaaggac atccttttgc agcagcagca gcagcttggc 300
cacggcggcc ccgaggcggc cccgcgcgcg ccgcaggcct tggagcgcta cccgttggcg 360
gccgcggccg agaggccccc gcgcctcggc tctgacttcg gcagcagccg cccggcagcg 420
agcctggccc agccgccgac gccgcagccg ccgcccgtga acggcatcct ggtgcccaac 480
ggcttctcca agctagagga gccgcccgag ctgaatcgcc agagcccgaa cccgcggcgc 540
ggccacgcgg tgccgcccac cctggtgccg ctcatgaacg gctcggccac gccgctgccc 600
accgcgctcg gcctcggcgg ccgcgctgcc gcctccttag ccgcggtgtc cggaaccgcg 660
gccgccagcc tgggctccgc gcagcccacc gatctgggcg cccacaagcg gccggcatcc 720
gtgtcgagca gcgctgccgt ggagcacgag cagcgtgagg cggcagccaa ggagaaacaa 780
ccgccgccgc ctgcgcaccg gggcccggcc gacagcctgt ccaccgcggc cggggccgcc 840
gagctgagcg cggaaggtgc gggcaagagc cgcgggtctg gagagcagga ctgggtcaac 900
aggcccaaga ccgtgcgcga cacgctgctg gcgctgcacc agcacggcca ctcggggccc 960
ttcgagagca agtttaagaa ggagccggcc ctgactgcag gcaggttgtt gggtttcgag 1020
gccaacgggg ccaacgggtc taaagcagtt gcaagaacag caaggaaaag gaagccctct 1080
ccagaaccag aaggtgaagt cgggcccccc aagatcaacg gagaggccca gccgtggctg 1140
tccacatcca cagaggggct caagatcccc atgactccta catcctcttt tgtgtctccg 1200
ccaccaccca ctgcctcacc tcattccaac cggaccacac cgcctgaagc ggcccagaat 1260
ggccagtccc ccatggcagc cctgatctta gtagcagaca atgcaggggg cagtcatgcc 1320
tcaaaagatg ccaaccaggt tcactccact accaggagga atagcaacag tccgccctct 1380
ccgtcctcta tgaaccaaag aaggctgggc cccagagagg tgggggggca gggagcaggc 1440
aacacaggag gactggagcc agtgcaccct gccagcctcc cggactcctc tctggcaacc 1500
agtgccccgc tgtgctgcac cctctgccac gagcggctgg aggacaccca ttttgtgcag 1560
tgcccgtccg tcccttcgca caagttctgc ttcccttgct ccagacaaag catcaaacag 1620
cagggagcta gtggagaggt ctattgtccc agtggggaaa aatgccctct tgtgggctcc 1680
aatgtccccc gggcctttat gcaaggggaa attgcaacca tccttgctgg agatgtgaaa 1740
gtgaaaaaag agagagactc gtgacttttc cggtttcaga aaaacccaat gattaccctt 1800
aattaaaact gcttgaattg tatatatatc tccatatata tatatatcca agacaaggga 1860

aatgtagact tcataaacat ggctgtataa ttttgatttt ttttgaatac attgtgtttc 1920
tatatttttt ttgacgacaa aaggtatgta cttataaaga catttttttc ttttgttaac 1980
gttattagca tatctttgtg ctttattatc ctggtgacag ttaccgttct atgtaggctg 2040
tgacttgcgc tgcttttttа gagcacttgg caaatcagaa atgcttctag ctgtatttgt 2100
atgcacttat tttaaaaaga aaaaaaaagc caaatacatt ttctgacatt gtaaaaaaaa 2160

<210> 33
<211> 2800
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7492451CB1

<400> 33
cgagctcgct cacgtattac ggcgcagtgt gctggcaaga ggatgtggga gatgccggcg 60
gctgctgctc aggtcctgca cgtctctgaa aaccccgtac ctctgagcgt cagagtgagc 120
cccgaggtcc gggacgggtg ggtgggtgcc tgccagcagc tgaggcgcca tggcggcggc 180
agcggtgtca gagtcttggc cggagctgga gctggctgag cgcgagcggc ggcgggagct 240
gctgctgacg gggcccgggc tggaggagcg agtgcgggcg gcgggtgggc agctgccgcc 300
gcggcttttc accctgccgc tgctgcacta cttggaagtg agcggctgcg ggagcttgcg 360
cgcgccgggg cctggcctgg cgcagggcct gccgcagctg cacagcctcg tgctgcggcg 420
caacgcgctg gggcccggcc tgagccccga gctcgggccg ctgcctgccc ttcgggtgct 480
cgacctgtcg ggcaacgcgc tggaggcgct gccgccgggc caaggcctgg gccccgccga 540
gccgccgggc cttccgcagc tgcagagcct caacctcagc ggcaaccggc tgcgcgagct 600
gccagccgac ctggcgcgct gcgccccgcg cctgcagagc ctcaacctca ccggcaattg 660
cctagactcc tttcccgccg agctctttcg ccccggcgcg ctgcccctgc tcagtgaact 720
ggcggctgct gacaactgcc tccgagaact cagccccgac atcgcccacc tggcctcgct 780
caagacgttg gacctctcga acaaccagct gagcgagatc cctgcagagc ttgcggactg 840
ccccaagctc aaggagatca atttccgtgg gaacaagctg agggacaagc gcctggagaa 900
gatggtcagc ggctgccaga ccagatccat cctggagtac ctgcgcgtcg gaggccgtgg 960
tggcgggaag ggcaagggcc gtgccgaggg ctcggagaag gaagagagcc ggaggaagag 1020
gagggagagg aagcagaggc gggaaggtgg tgatggggag gagcaggacg tgggagatgc 1080
cggccggctg ctgctcaggg tcctgcacgt ctctgaaaac cccgtacctc tgacagtcag 1140
agtgagcccc gaggtccggg atgtgcggcc ctacattgtg ggggccgtgg tgcgaggcat 1200
ggacctgcag ccagggaatg cactcaagcg cttcctcacc tcgcagacca agctccacga 1260
agatctctgt gagaagagga cggctgccac ccttgccacc cacgagctcc gtgccgtcaa 1320
agggcccctg ctgtactgcg cccggccccc acaggacctc aagattgtcc ccttggggcg 1380
gaaagaagac aaggccaagg agctggtgcg gcagctgcag ctggaggccg aggagcagag 1440
gaagcagaag aagcggcaga gtgtgtcggg cctgcacaga taccttcact tgctggatgg 1500
aaatgaaaat tacccgtgtc ttgtggatgc agacggtgat gtgatttcct tcccaccaat 1560
aaccaacagt gagaagacaa aggttaagaa aacgacttct gatttgtttt tggaagtaac 1620
aagtgccacc agtctgcaga tttgcaagga tgtcatggat gccctcattc tgaaaatggc 1680
agaaatgaaa aagtacactt tagaaaataa agaggaagga tcactctcag atactgaagc 1740
cgatgcagtc tctggacaac ttccagatcc cacaacgaat cccagtgctg gaaaggacgg 1800
gccctccctt ctggtggtgg agcaggtccg ggtggtggat ctggaaggga gcctgaaggt 1860
ggtgtacccg tccaaggccg acctggccac tgcccctccc cacgtgactg tcgtgcgctg 1920
acgccagggc cgcctgtccg cgtttgtttg gccggttttg cggaggtttc tatgcggcaa 1980
tgctgaatta tccgttagat tttcacccca gtttttttgt tggttttttt tttttgagat 2040
ggagtctcgc tctgtcgcca ggctggagtg cagtggcgtg atctcgagtc actgcagcct 2100
gtgtctcctg ggttcaagcg attctcctgc ctcagcctcc caagtagctg ggactacagg 2160
tgtgtgccac taagctcagc taatttttgt atttttagta gagacggggt ttcaccattt 2220
tgaccaggat ggtcttgatc tcttgacctc atgatctgtc cacctctgcc cctcaaagtg 2280
ctgggattac gtgatccacc cgcctcagcc tcccaaagtg ctgggattac aggcgtgagc 2340
tgtgcctggc ccaccccagc atttttttaa gatgtatgta ttcgttgttc tgtttttcca 2400

gatgattctg tcgtaaagtg atgctatgtt gtcgttacaa catcaaagtg attttacggt 2460
ttttgatggg attattcaag tgtcagaatt aactgttcaa aatgttctga atcatgtaga 2520
tacatggcag gtaactgttt atgggagaaa agtacagtgc tgttacgtgg cactgtacag 2580
tcatgtgcca cgtaacagcg tctgggtcag tgacggacac ttacctgaca gcggatccac 2640
aatattctcg tgcagtgtgt ttggaatcct ggtctgggct ctcggcgttg gccttgtaga 2700
tcaagtaggg gaagtgagtg atgttcagtc atgctgctgg gacacttggt tttccagatg 2760
aaaacacata aataaaacta catgcaccat caaaaaaaaa 2800

<210> 34
<211> 1384
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4650669CB1

<400> 34
gagggaggcg ggaccggccg cgtctccatg gcgacgcggg acgcggcggt ggaggcggtt 60
gctacccacc tgctagaggc gctgcggctg tgatccaggc tggggcgaga ccatgtcgga 120
cctgggctcg gaggagttgg aggaggaggg agagaatgat attggggaat atgaggggg 180
tcggaatgag gcaggcgaaa ggcacggacg tgggagggca cggctaccca acggggacac 240
ctacgaaggg agctacgaat tcggtaaaag acatggccag gggatctaca aatttaaaaa 300
tggtgctcga tatatcggag aatatgttag aaataaaaag cacggtcaag gcacttttat 360
atatccagat ggatccagat atgaaggaga gtgggcaaat gacctgcggc acggccatgg 420
cgtatactac tacatcaata atgacaccta cactggagag tggtttgctc atcaaaggca 480
tgggcaaggc acctatttat acgcggagac gggcagtaag tatgttggca cctgggtgaa 540
cggacagcag gagggcacgg ccgagctcat tcacctgaac cacaggtacc agggcaagtt 600
cttgaacaaa aatcctgttg gccctggaaa gtatgtattt gatgttgggt gtgaacaaca 660
tggtgaatat cgtttaacag atatggaaag aggagaagag gaagaggagg aagaattagt 720
aactgttgtt ccaaaatgga aagctaccca aatcactgaa ttggccctgt ggacaccaac 780
tctccccaaa aagccgacct ctacggatgg acctggccaa gacgctccag gagctgagag 840
tgcaggagaa cccggggagg aggcccaggc tctgctggag ggcttcgagg gtgagatgga 900
catgaggcct ggagatgaag atgcagacgt cctccgggaa gagagccggg agtatgacca 960
ggaggagttc cgctatgaca tggatgaggg aaacattaat tctgaagaag aagaaactag 1020
acagtcagac ctccaggact aagatgaagt gagccgagag gagatcgtat cataagaatg 1080
cttctgtcgt tagccgggtg cagtgctgtg tgtatctagt tccagctact tgagaggctg 1140
aggcaggagg attgcttgag tccagaaagt ggcagttgca gtgagtggag atcgcgccac 1200
tgctctccag cctgggtggc agagcgagac cctgtctcaa aaaataaaca aaaacaaaat 1260
gcttctgtca gttaacaatc tttattagag ggtttttagt ctttctttct cagctgtatg 1320
ttaagttggt tgacaaatgc aaataaacgt ctttattatc ctttctttct gaaaaaaaaa 1380
aaaa 1384

<210> 35
<211> 969
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7485268CB1

<400> 35
tggtgcggcc cagagtgtag cagcccctga gcctcggcct tggcctcacc cactgggcca 60
ggaccgtccg tgcccaggca gaagctcccg cctccactgg ctgtattccc aacgggggca 120
gggccgggct gccgtagggt ctgggggctc cagctggggc acctccttct cagtccctcc 180

cctctttcct ctcccccagc cctggccaag atggcagccc ccgccctgct gctcctagca 240
ctgctgctgc ccgtgggggc ctggcccggg ctgcccagga ggccctgtgt gcactgctgc 300
cgcccggcct ggcccccctgg accctatgcc cgggtgagtg acagggacct gtggaggggg 360
gacctgtgga gggggctgcc tcgagtacgg cccactatag acatcgaaat cctcaaaggt 420
gagaagggtg aggccggcgt ccgaggtcgg gccggcagga gcgggaaaga ggggccgcca 480
ggcgcccggg gcctgcaggg ccgcagaggc cagaaggggc aggtggggcc gccgggcgcc 540
gcgtgccgac gtgcctacgc cgccttctcc gtgggccggc gcgagggcct gcacagctcc 600
gaccacttcc aggcggtgcc cttcgacacg gagctggtga acctggacgg cgccttcgac 660
ctggccgcgg gccgcttcct ctgcacggtg cccggcgtct acttcctcag cctcaacgtg 720
cacacctgga actacaagga gacctacctg cacatcatgc tgaaccggcg gcccgcggcc 780
gtgctctacg cgcagcccag cgagcgcagc gtcatgcagg cccagagcct gatgctgctg 840
ctggcggcgg gcgacgccgt ctgggtgcgc atgttccagc gcgaccggga caacgccatc 900
tacggcgagc acggagacct ctacatcacc ttcagcggcc acctggtcaa gccggccgcc 960
gagctgtag 969

<210> 36
<211> 2792
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2112995CB1

<400> 36
cacacaacca gccacccccct ctaggatccc agcccagctg gtgctgggct cagaggagaa 60
ggccccgtgt tgggagcacc ctgcttgcct ggagggacaa gtttccggga gagatcaata 120
aaggaaagga aagagacaag gaagggagag gtcaggagag cgcttgattg gaggagaagg 180
gccagagaat gtcgtcccag ccagcaggga accagacctc ccccggggcc acagaggact 240
actcctatgg cagctggtac atcgatgagc cccagggggg cgaggagctc cagccagagg 300
gggaagtgcc ctcctgccac accagcatac cacccggcct gtaccacgcc tgcctggcct 360
cgctgtcaat ccttgtgctg ctgctcctgg ccatgctggt gaggcgccgc cagctctggc 420
ctgactgtgt gcgtggcagg cccggcctgc ccagccctgt ggatttcttg gctggggaca 480
ggccccgggc agtgcctgct gctgttttca tggtcctcct gagctccctg tgtttgctgc 540
tccccgacga ggacgcattg cccttcctga ctctcgcctc agcacccagc caagatggga 600
aaactgaggc tccaagaggg gcctggaaga tactgggact gttctattat gctgccctct 660
actaccctct ggctgcctgt gccacggctg gccacacagc tgcacacctg ctcggcagca 720
cgctgtcctg ggcccacctt ggggtccagg tctggcagag ggcagagtgt ccccaggtgc 780
ccaagatcta caagtactac tccctgctgg cctccctgcc tctcctgctg ggcctcggat 840
tcctgagcct ttggtaccct gtgcagctgg tgagaagctt cagccgtagg acaggagcag 900
gctccaaggg gctgcagagc agctactctg aggaatatct gaggaacctc ctttgcagga 960
agaagctggg aagcagctac cacacctcca agcatggctt cctgtcctgg gcccgcgtct 1020
gcttgagaca ctgcatctac actccacagc caggattcca tctcccgctg aagctggtgc 1080
tttcagctac actgacaggg acggccattt accaggtggc cctgctgctg ctggtgggcg 1140
tggtacccac tatccagaag gtgagggcag gggtcaccac ggatgtctcc tacctgctgg 1200
ccggctttgg aatcgtgctc tccgaggaca agcaggaggt ggtggagctg gtgaagcacc 1260
atctgtgggc tctggaagtg tgctacatct cagccttggt cttgtcctgc ttactcacct 1320
tcctggtcct gatgcgctca ctggtgacac acaggaccaa ccttcgagct ctgcaccgag 1380
gagctgccct ggacttgagt cccttgcatc ggagtcccca tccctcccgc caagccatat 1440
tctgttggat gagcttcagt gcctaccaga cagcctttat ctgccttggg ctcctggtgc 1500
agcagatcat cttcttcctg ggaaccacgg ccctggcctt cctggtgctc atgcctgtgc 1560
tccatggcag gaacctcctg ctcttccgtt ccctggagtc ctcgtggccc ttctggctga 1620
ctttggccct ggctgtgatc ctgcagaaca tggcagccca ttgggtcttc ctggagactc 1680
atgatggaca cccacagctg accaaccggc gagtgctcta tgcagccacc tttcttctct 1740
tcccccctcaa tgtgctggtg ggtgccatgg tggccacctg gcgagtgctc ctctctgccc 1800
tctacaacgc catccacctt ggccagatgg acctcagcct gctgccaccg agagccgcca 1860

ctctcgaccc cggctactac acgtaccgaa acttcttgaa gattgaagtc agccagtcgc 1920
atccagccat gacagccttc tgctccctgc tcctgcaagc gcagagcctc ctacccagga 1980
ccatggcagc cccccaggac agcctcagac caggggagga agacgaaggg atgcagctgc 2040
tacagacaaa ggactccatg gccaagggag ctaggcccgg ggccagccgc ggcagggctc 2100
gctggggtct ggcctacacg ctgctgcaca acccaaccct gcaggtcttc cgcaagacgg 2160
ccctgttggg tgccaatggt gcccagccct gagggcaggg aaggtcaacc cacctgccca 2220
tctgtgctga ggcatgttcc tgcctaccat cctcctccct ccccggctct cctcccagca 2280
tcacaccagc catgcagcca gcaggtcctc cggatcactg tggttgggtg gaggtctgtc 2340
tgcactggga gcctcaggag ggctctgctc cacccacttg gctatgggag agccagcagg 2400
ggttctggag aaaaaaactg gtgggttagg gccttggtcc aggagccagt tgagccaggg 2460
cagccacatc caggcgtctc cctaccctgg ctctgccatc agccttgaag ggcctcgatg 2520
aagccttctc tggaaccact ccagcccagc tccacctcag ccttggcctt cacgctgtgg 2580
aagcagccaa ggcacttcct caccccctca gcgccacgga cctctctggg gagtggccgg 2640
aaagctcccg ggcctctggc ctgcagggca gcccaagtca tgactcagac caggtcccac 2700
actgagctgc ccacactcga gagccagata tttttgtagt ttttatgcct ttggctatta 2760
tgaaagaggt tagtgtgttc cctgcaataa ac 2792

<210> 37
<211> 3567
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1613452CB1

<400> 37
gtggtccctg cctggctgag gtggcagcag gggggcggga cgcgcagcta tggcagaggg 60
cagcggggaa gtggtcacag tgtctgcgac cggggctgcc aacggcctca acaatggggc 120
aggcgggacc tcggcgacga ccagcaaccc gctgtcgcgc aagctgcata agatcctgga 180
gacgcggctg gacaacgaca aggagatgtt agaagctctc aaggcacttt caacctttt 240
tgttgaaaat agtctgcgga ctcgaagaaa tttacgtgga gatattgaac gtaaaagttt 300
agccatcaat gaagaatttg taagcatttt caaggaagtg aaggaggaac ttgaaagcat 360
aagcgaagat gttcaagcaa tgagcaactg ttgtcaagat atgacaagtc gcctacaggc 420
agcaaaggaa cagactcaag atttaatagt taaaaccact aagcttcaat ctgaaagcca 480
aaaattagag ataagagctc aagttgcaga tgccttctta tccaagttcc aactgacttc 540
tgatgaaatg agtcttcttc gaggtacaag agaaggaccc attactgagg atttttcaa 600
ggcactggga agagtaaaac agattcataa tgatgtcaaa gttctcttgc gtacaaatca 660
acaaacggca ggtttagaaa ttatggaaca gatggcctta cttcaagaaa cggcttatga 720
aagactttac cgatgggctc aaagtgaatg cagaacattg acacaagaat catgtgacgt 780
atctccagta ttgacacagg caatggaagc cctgcaggac agacctgtct tatataaata 840
taccttagat gaatttggaa cagccagaag aagtacagtt gttcgtggat ttattgatgc 900
gctcacaaga gggggccccg gaggtacacc tagaccaatt gaaatgcatt ctcatgaccc 960
tttgaggtat gtaggagata tgttggcttg gctccatcaa gctactgctt ctgaaaagga 1020
acaccttgaa gctctcttaa agcatgtaac tacacaaggt gttgaagaaa atattcaaga 1080
agttgttggg catatcactg aaggtgtgtg caggcctcta aaggttcgaa ttgagcaagt 1140
aatagttgct gaacctgggg cagttttatt atataaaatt tctaatctcc tcaaatttta 1200
tcaccataca atcagtggta ttgttggaaa tagtgcaact gcattattga ctaccattga 1260
agaaatgcat ttgctaagca aaaaaatatt cttcaatagc ttgagtcttc atgcaagtaa 1320
attaatggac aaggttgaac tcccaccacc tgatcttgga ccaagttctg cactaaatca 1380
gacactcatg ttgctgcgtg aagttttagc atctcacgat tcttcagttg taccattaga 1440
tgctcgtcaa gctgattttg tgcaggtttt atcatgtgtc ttggatcctc tcctacagat 1500
gtgtactgta tcagccagca atttaggcac agctgacatg gccactttca tggtcaattc 1560
actatatatg atgaagacaa cattagctct atttgaattc actgacagac gtctggaaat 1620
gctacagttt cagatcgaag cacatttgga cacacttata aatgagcaag cctcttatgt 1680
tttaactagg gtaggcttga gttacatcta taacactgta cagcaacata aacctgaaca 1740

gggctcttta gctaatatgc ccaacctaga ttctgtgaca ctgaaggctg caatggttca 1800
gtttgatcgt tatctgtcag ccccagacaa cctattgata ccacagctga actttcttct 1860
aagtgccaca gtgaaagagc agatcgtaaa acaatctaca gaattagtct gcagagccta 1920
tggtgaagtg tatgcagccg tgatgaatcc aatcaatgaa tacaaagatc cagagaacat 1980
tcttcaccga tcgccgcagc aagtgcagac gcttctttcc tgattatctt atttcattgt 2040
gttagcaaaa tatgacctcc ctaaaacact gaaggttatt ttttattctt tgaattttta 2100
ctttataatt tgatagttac agttttcttt gtatcataag attgtaagtc ccgataattt 2160
ttttttttt ggtctcagta acagggaagt aagtaacatg ttgacctgag ctagtattgc 2220
tgtgtatcta ctctaaatga gatgatctat ttttttgcta gccatctctc cagctctgca 2280
gttttcactg tattcaggaa gcataaagta gtatgaaagg tttgaagaat tttttttac 2340
aagactagtt ctaaattaac agcttataaa aaatttgtct aaatttaata attagtataa 2400
ggatatgacc taataaatgt ctccttacct aaagattcat ttgctttctt ttaatatgag 2460
taggcatact tagtagcttt tctgaaccta gcctatgtct ctgtccccaa aatagctgcc 2520
cttaaagagt tgttagcaga gagaaaaata acagtgaatg tgctcctggt gtatatggca 2580
gtgaatctcc tttctgttct actttagcat actatatata tttgactgtg tacattctta 2640
tgcaatttta agtatacact cagcaataat tagaaaaaaa ggagagagaa aagtgattta 2700
aacagggtgg attccactct gtgggagcct tcgatggaac tcaaggtgga gctcagcctt 2760
tccaatgagc tctaagcatg tagatagcct gagctgtgtc taagcctggt gtttaaagat 2820
gggtatttgt catacaatat gggtcctaaa tccaaccaac tacacatttt atctggtgtt 2880
caaaccaaag aaacaatgat ctactcaaac attggagaaa aaaactgcca gaggaggagt 2940
tgccaattgg cagtgtgtct tatctccatg ttgtaactgg actctgactt tagaccatta 3000
cctattagga agattaaaaa tgactgtatt tttaaaggaa taaatcccag tgtgcctgat 3060
ttgacattct tgtcagcaaa aaaaaactta atttctagta aatctataaa aatgggtaag 3120
tccctaaatt acaaatgaga aaattgaagc acaaggaaaa aaataactag tttgaaatat 3180
tttgaaaagt aataacataa aactagtatt tgtagaagat tatgtgttgt atataacaaa 3240
ttagtattta tagaatatga cctatttatc tgaagtttat aattgtttat acctaataca 3300
gttctttttg gagtaagaat gattatataa tcgttatcca tttgggtata aatctgtatt 3360
tttagttttt tccctttgat tagtatgtgt tacatataaa gacagaaaat aaagtataaa 3420
tctagagctt aaattgtata taatttattt ctacagagaa agaagattga taccttgcta 3480
tgagtgaatt cctttgtttt atagggaaaa tttattgtgc tttttacctg gttttttcaa 3540
ataaaatatt aaaatattaa aaaaaaa 3567

<210> 38
<211> 6004
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55061615CB1

<400> 38
actgctttcg tccggcttgg ggccccctca agctggggca agggcttgga ctgctctcaa 60
gttgttctga ttggaattca gtcacagact caggtacttg ctgtccccca acgtcccgtg 120
aactacaatg acaatcctga atgcattact ccatgcagac ccagtccagg gtaagcgaat 180
tcagctgaaa gccagggcat tcgaactctc cgaaggagat gtcctcaagg tttatgatgg 240
caacaacaac tccgcccgtt tgctgggagt ttttagccat tctgagatga tgggggtgac 300
tttgaacagc acatccagca gtctgtggct tgatttcatc actgatgctg aaaacaccag 360
caagggcttt gaactgcact tttccagctt tgaactcatc aaatgtgagg acccaggaac 420
ccccaagttt ggctacaagg ttcatgatga aggtcatttt gcagggagct ccgtgtcctt 480
cagctgtgac cctggataca gcctgcgggg tagtgaggag ctgctgtgtc tgagtggaga 540
gcgccggacc tgggaccggc ctctgcccac ctgtgtcgcc gagtgtggag ggacagtgag 600
aggagaggtg tcgggcagg tgctgtcacc cgggtatcca gctccctatg aacacaatct 660
caactgcatc tggaccatcg aagcagaggc cggctgcacc attgggctac acttcctggt 720
gtttgacaca gaggaggttc acgacgtgct gcgcatctgg gatgggcctg tggagagcgg 780
ggttctgctg aaggagctga gtggcccggc cctgcccaag gacctgcata gcaccttcaa 840

ctcggtcgtc ctgcagttca gcactgactt cttcaccagc aagcagggct ttgccattca 900
attttcagtg tccacagcaa cgtcctgcaa tgaccctggg atcccgcaga atgggagtcg 960
gagtggtgac agttgggaag ccggcgactc cacagtgttc cagtgtgacc ctggctacgc 1020
gctgcaggga agtgcagaga tcagctgtgt gaagatcgag aacaggttct tctggcagcc 1080
cagcccgcca acatgcatcg ctccctgcgg gggagacctg acaggaccat ctggagtcat 1140
cctctcacca aattacccag aaccctaccc gccaggcaag gagtgtgact ggaaagtgac 1200
cgtctcacca gactacgtca tcgccctggt atttaacatc tttaacctgg agcctggcta 1260
tgacttcctc catatctacg acggacggga ctctctcagc cctctcatag gaagcttcta 1320
tggctcccag ctcccaggcc gcattgaaag cagcagcaac agcctcttcc tcgccttccg 1380
cagcgatgca tctgtgagca atgctggctt cgtcattgac tatacagaaa acccgcggga 1440
gtcatgtttt gatcctggtt ccatcaagag cggcacacgg gtggggtccg acctgaagct 1500
gggctcctcc gtcacctact actgccacgg gggctacgaa gttgagggca cctcgaccct 1560
gagctgcatc ctggggcctg atgggaagcc cgtgtggaac aatccccggc cagtctgcac 1620
agccccctgt gggggacagt atgtgggttc ggacggagtg gtcttgtccc ccaactaccc 1680
ccagaactac accagtggac agatctgctt gtattttgtt actgtgccca aggactatgt 1740
ggtgtttggc cagttcgcct tctttcacac ggccctcaac gacgtggtgg aggttcacga 1800
cggccacagc cagcactcgc ggctcctcag ctccctctcg ggctcccata caggtatccg 1860
gggctcggcc agtgtgggga tggttgtggg ccgggggcat cacgtccggc taaaggaagg 1920
aggctctaga agcaccccat ggccgcaggt ggaaccctac ggctctgcgt gcctgtcgtg 1980
ttctggtgct tgtctacaac gcagcagcca gctcgtgaga gctccaacta gcggggcctt 2040
cagcagctgc cctcacccag actgtgtcta caccgccccc ttgtggtgta gccttctcct 2100
gttgaatggc aactacacta attggctgca ggtccagttg gtgctgtctc tcccctggcc 2160
catctgtact gcaccaagca gaagatatac ctttgtcttc tgctacaaaa gctgtcagtc 2220
taccctggtt tcctgtgccc atgcaggaga atcactgccc ttggccacct ccaatcaagt 2280
tctcattaag ttcagcgcca aaggcctcgc accagccaga ggcttccact ttgtctacca 2340
aggtatggag gacatggacg ccggagcggt tcctcgaacc agcgccacgc agtgcagctc 2400
tgtgccggaa ccccgctatg gcaagaggct gggcagtgac ttctcggtgg gggccatcgt 2460
ccgcttcgaa tgcaactccg gctatgccct gcaggggtcg ccagagatcg agtgcctccc 2520
tgtgcctggg gccttggccc aatggaatgt ctcagcgccc acgtgtgtgg tgccgtgtgg 2580
aggcaacctc acagagcgca ggggcaccat cctgtccccc ggcttcccag agccgtacct 2640
caacagcctc aactgtgtgt ggaagatcgt ggtccccgaa ggcgctggca tccagatcca 2700
agttgtcagt tttgtgacag agcagaactg ggactcgctg gaagtatttg atggtgcaga 2760
taacactgta accatgctgg ggagtttctc aggaacaacc gtgcctgccc ttctgaacag 2820
cacctccaac cagctctacc ttcatttcta ctcagatatc agcgtatctg cagctggctt 2880
ccacttggag tacaaaacgg tgggcctgag cagttgtccg gaacctgctg tgcccagtaa 2940
cggggtgaag actggcgagc gctacttggt gaatgatgtg gtgtctttcc agtgtgagcc 3000
gggatatgcc ctccagggcc acgcccacat ctcctgcatg cccggaacag tgcggcgatg 3060
gaactaccct cctccactct gtattgcaca gtgtgggggа acagtggagg agatggaggg 3120
ggtgatcctg agccccggct tcccaggcaa ctaccccagt aacatggact gctcctggaa 3180
aatagcactg cccgtgggct ttggagctca catccagttc ctgaacttct ccaccgagcc 3240
caaccacgac tacatagaaa tccggaatgg cccctatgag accagccgca tgatgggaag 3300
attcagtgga agcgagcttc caagctccct cctctccacg tcccacgaga ccaccgtgta 3360
tttccacagc gaccactccc agaatcggcc aggattcaag ctggagtatc aggatttgac 3420
ttactcccac cagatttctt ccttcctgag aggttttgat ctctcggagt tggaaagaac 3480
caactcaact cctcccgtcg ccgcttccta tgtctgggat cttgatcctg gttgtgaagc 3540
ctatgaactt caagagtgcc cagacccaga gcccttтgcc aatggcattg tgaggggagc 3600
tggctacaac gtgggacaat cagtgacctt cgagtgcctc ccggggtatc aattgactgg 3660
ccaccctgtc ctcacgtgtc aacatggcac caaccggaac tgggaccacc ccctgcccaa 3720
gtgtgaagtc ccttgtggcg ggaacatcac ttcttccaac ggcactgtgt actccccggg 3780
gttccctagc ccgtactcca gctcccagga ctgtgtctgg ctgatcaccg tggcccaatt 3840
ggccatgggc gtccgcctca acctcagcct gctgcagaca gagccctctg gagatttcat 3900
caccatctgg gatgggccac agcaaacagc accacggctc ggcgtcttca cccggagcat 3960
ggccaagaaa acagtgcaga gttcatccaa ccaggtcctg ctcaagttcc accgtgatgc 4020
agccacaggg gggatcttcg ccatagcttt ctccgcttat ccactcacca aatgccctcc 4080
tcccaccatc ctccccaacg ccgaagtcgt cacagagaat gaagaattca atataggtga 4140
catcgtacgc tacagatgcc tccctggctt taccttagtg gggaatgaaa ttctgacctg 4200

caaacttgga acctacctgc agtttgaagg accacccccg atatgtgaag tgcactgtcc 4260
aacaaatgag cttctgacag actccacagg cgtgatcctg agccagagct accctggaag 4320
ctatccccag ttccagacct gctcttggct ggtgagagtg gagcccgact ataacatctc 4380
cctcacagtg gagtacttcc tcagcgagaa gcaatatgat gagtttgaga tttttgatgg 4440
tccatcagga cagagtcctc tgctgaaagc cctcagtggg aattactcag ctcccctgat 4500
tgtcaccagc tcaagcaact ctgtgtacct gcgttggtca tctgatcacg cctacaatcg 4560
gaagggcttc aagatccgct attcaggcca gaccagcacc cagcccgggg gctccatcca 4620
ctttggctgc aacgccggct accgcctggt gggacacagc atggccatct gtacccggca 4680
cccccagggc taccacctgt ggagcgaagc catccctctc tgtcaagctc tttcctgtgg 4740
gcttcctgag gcccccaaga atggaatggt gtttggcaag gagtacacag tgggaaccaa 4800
ggccatgtac agctgcagtg aaggctacca cctccaggca ggcgctgagg ccactgcaga 4860
gtgtctggac acaggcctat ggagcaaccg caatgtccca ccacagtgtg tccgtgagtc 4920
ctcgggcaat ggaggcgggt ctgtgacttg tcctgatgtc agtagcatca gcgtggagca 4980
tggccgatgg aggcttatct ttgagacaca gtatcagttc caggcccagc tgatgctcat 5040
ctgtgaccct ggctactact atactggcca aagggtcatc cgctgtcagg ccaatggcaa 5100
atggagcctc ggggactcta cgcccacctg ccgaatcatc tcctgtggag agctcccgat 5160
tccccccaat ggccaccgca tcggaacact gtctgtctac ggggcaacag ccatcttctc 5220
ctgcaattcc ggatacacac tggtgggctc cagggtgcgt gagtgcatgg ccaatgggct 5280
ctggagtggc tctgaagtcc gctgccttgc cactcagacc aagctccact ccattttcta 5340
taagctcctc ttcgatgtac tctcttcccc atccctcacc aaagctggac actgtgggac 5400
tcctgagccc attgtcaacg gacacatcaa tggggagaac tacagctacc ggggcagtgt 5460
ggtgtaccaa tgcaatgctg gcttccgcct gatcggcatg tctgtgcgca tctgccagca 5520
ggatcatcac tggtcgggca agaccccttt ctgtgtgcat gttaagcagc agttgctgct 5580
gctgctgctg ctgttgtgtg atgatgatga tgatgaagat gatggtagtg gtgcaattac 5640
ctgtggacac ccaggcaacc ctgtcaacgg cctcactcag ggtaaccagt ttaacctcaa 5700
cgatgtggtc aagtttgttt gcaaccctgg gtatatggct gaggggggctg ctaggtccca 5760
atgcctggcc agcgggcaat ggagtgacat gctgcccacc tgcagaatca tcaactgtac 5820
agatcctgga caccaagaaa atagtgttcg tcaggtccac gccagcggcc cgcacaggtt 5880
cagcttcggc accactgtgt cttaccggtg caaccacggc ttctacctcc tgggcacccc 5940
agtgctcagc tgccagggag atggcacatg ggaccgtccc cgcccccagt gtctctgtaa 6000
gtag 6004

<210> 39
<211> 1917
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7503435CB1

<220>
<221> unsure
<222> 1483-1517, 1556-1589
<223> a, t, c, g, or other

<400> 39
ctcaaaggca aacaaaagga aatgcccggc tccccatggc tgtggccagc accttcatac 60
cagggctcaa ccctcagaac cctcattata tcccagggta agactcccac ttagcgctcc 120
ccgcttgctg ctgggagtag gggtataggg tggaggaatt aagctgctta gagcatgggg 180
ccaagcggac aggttggttt gggaggctga ggtctctatg gtggtggatg gaaaaggaga 240
agcaagggac tagtgagaca ccgggtctgc cccaaagctt tgtgttgttc tctccatggg 300
aacagggtaa cagcagagga aggatcgaaa gcccaggcag agaacttaga aatgcctgct 360
ccaaatatat acactgctga gtccagaaag atcagggctt cagaggtctc cgcttctgcc 420
tctggggatg gcggggcagg gcgattctgg tgagtgggag agtctgtgct gcaggtacac 480
tggacactgc ccactacttc ggttcagcgt gggccagacc tatgggcagg tgactggtca 540

gctacttcga ggccctcctg gcctagcctg gccccctgtc caccgcacac ttctgcctcc 600
cattcggcct ccaagatctc ctgaggttcc cagggagagt ctacctgtca ggcgtgggca 660
agaaaggctc agctccagca tgatccctgg gtacacaggt ttgtacgcag gtatgcacag 720
gtgcccctcc caggagcatg tgttccaaac actaacgagt cttccttgtc cctgcctgcc 780
caggttttgt accccgggca cagttcatct ttgccaagaa ctgcagccag gtctgggccg 840
aggctctgag tgactttact cacttgcatg aaaagcaagg gagtgaagag ctaccaaagg 900
aggccaaggg aagaaaggac acagagaagg accaggtgcc agagccggag gggcagctgg 960
aggagccgac actggaggtg gtggaacaag cttctcccta ctccatggat gacagggacc 1020
ctcggaagtt cttcatgtca gctttcctgt gctcaccaac caggcactgc aggaatttgg 1080
gcagaagcac tcaccaggca gtgcccagga ccccaaacat ctccccccac ttcccagaac 1140
ataccctcag aacctgggtc ttttacctaa ctatgggggc tacgtgccag ggtataagtt 1200
ccagtttggc cacacatttg gccatctcac ccatgatgct ctgggcctca gcaccttcca 1260
gaagcagctc ttggcttagg ccactggaca tcaagttccc ttccctttc atcctatccc 1320
agccatcctt ttggaaggga gagaggtggg tgggagggtg ggagggtggg ggaacacaaa 1380
gagaaaatgg tttggaggct gagcaccttt tttattaata ggtataataa ataaataaat 1440
aaatacataa acagaaaaaa aaaaaaaagg ggcgggccgg cgnnnnnnnn nnnnnnnnnn 1500
nnnnnnnnnn nnnnnnngac ggtcataacg gcctggcagg gtaccgggtc cgggannnnn 1560
nnnnnnnnnn nnnnnnnnnn nnnnnnnnng gggccggccc ttttttttt ttccctttt 1620
tttagtttat ttttattatt attgtttttt ctctattata tgatcgaaga ttttgtaatg 1680
tggtggtcta tatgtgcact cccgagagat tattacaaaa cgacggagga cacacaacag 1740
acgggggtgc gcaaccacgc ggtgggtcaa accgccggga aaaaaaccac ccccctgggg 1800
gagaagagac accaagtacc acagtagaga gagaagccag gggggcaata cgcgaccaag 1860
agggagaagg gccacaagag gacgcacacg aagagtggac tctagagcaa caagaac 1917

<210> 40
<211> 1208
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 7504149CB1

<400> 40
ggcggttgct acccacctgc tagaggcgct gcggctgtga tccaggctgg ggcgagacca 60
tgtcggacct gggctcggag gagttggagg aggagggaga gaatgatatt ggggggatct 120
acaaatttaa aaatggtgct cgatatatcg gagaatatgt tagaaataaa aagcacggtc 180
aaggcacttt tatatatcca gatggatcca gatatgaagg agagtgggca aatgacctgc 240
ggcacggcca tggcgtatac tactacatca ataatgacac ctacactgga gagtggtttg 300
ctcatcaaag gcatgggcaa ggcacctatt tatacgcgga gacgggcagt aagtatgttg 360
gcacctgggt gaacggacag caggagggca cggccgagct cattcacctg aaccacaggt 420
accagggcaa gttcttgaac aaaaatcctg ttggccctgg aaagtatgta tttgatgttg 480
ggtgtgaaca acatggtgaa tatcgtttaa cagatatgga aagaggagaa gaggaagagg 540
aggaagaatt agtaactgtt gttccaaaat ggaaagctac ccaaatcact gaattggccc 600
tgtggacacc aactctcccc aaaaagccga cctctacgga tggacctggc caagacgctc 660
caggagctga gagtgcagga gaacccgggg aggaggccca ggctctgctg gagggcttcg 720
agggtgagat ggacatgagg cctggagatg aagatgcaga cgtcctccgg gaagagagcc 780
gggagtatga ccaggaggag ttccgctatg acatggatga gggaaacatt aattctgaag 840
aagaagaaac tagacagtca gacctccagg actaagatga agtgagccga gaggagatcg 900
tatcataaga atgcttctgt cgttagccgg gtgcagtgct gtgtgtatct agttccagct 960
acttgagagg ctgaggcagg aggattgctt gagtccagaa agtggcagtt gcagtgagtg 1020
gagatcgcgc cactgctctc cagcctgggt ggcagagcga gaccctgtct caaaaaaata 1080
aacaaaaaca aaatgcttct gtcagttaac aatctttatt agagggtttt tagtctttct 1140
ttctcagctg tatgttaagt tggttgacaa atgcaaataa acgtctttat tatcctttct 1200
ttctgaaa 1208

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 September 2002 (12.09.2002)

PCT

(10) International Publication Number
WO 02/070709 A3

(51) International Patent Classification[7]: C12N 15/12, 5/10, C12Q 1/68, C07K 14/47, 16/18, G01N 33/50, 33/68, A01K 67/027, A61K 38/17, 39/395

(21) International Application Number: PCT/US02/03709

(22) International Filing Date: 8 February 2002 (08.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/268,117	9 February 2001 (09.02.2001)	US
60/269,618	15 February 2001 (15.02.2001)	US
60/271,118	23 February 2001 (23.02.2001)	US
60/274,486	7 March 2001 (07.03.2001)	US
60/274,436	9 March 2001 (09.03.2001)	US
60/334,229	28 November 2001 (28.11.2001)	US
60/353,284	1 February 2002 (01.02.2002)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandero, CA 94577 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 40633 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street, #205, San Leandro, CA 94577 (US). **ELLIOT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **HONCHELL, Cynthia, D.** [US/US]; 400 Laurel Street, #203, San Carlos, CA 94070 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **DING, Li** [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94506 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **GHANDI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue, #23, Mountain View, CA 94043 (US). **SANJANWALA, Madhusudan, M.** [US/US]; 210 Silvia Court, Los Altos, CA 94024 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **SWARNAKER, Anita** [CA/US]; 8 Locksley Avenue, #5D, San Francisco, CA 94122 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **SAPPERSTEIN, Stephanie, K.** [US/US]; 3531 Highland Avenue, Redwood City, CA 94062 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **MARQUIS, Joseph, P.** [US/US]; 3299 Moulin Lane, San Jose, CA 95135 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
27 February 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: MOLECULES FOR DISEASE DETECTION AND TREATMENT

(57) Abstract: The invention provides full-length human molecules for disease detection and treatment (MDDT) and polynucleotides which identify and encode MDDT. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of MDDT.

WO 02/070709 A3

INTERNATIONAL SEARCH REPORT

International Application No
PCT/[redacted] 02/03709

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C12N5/10 C12Q1/68 C07K14/47 C07K16/18
G01N33/50 G01N33/68 A01K67/027 A61K38/17 A61K39/395

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00 28027 A (INCYTE PHARMA INC ;KLINGLER TOD M (US); VOLKMUTH WAYNE (US); WALKE) 18 May 2000 (2000-05-18) See SEQ ID NO:3 - 100% identical over most of its length to SEQ ID NO:21 of the present application. the whole document -----	1-20,22, 23, 25-56,76

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 August 2002	14. 11. 02

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Sprinks, M

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 02/03709

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 19, 22, 33 and 35 are directed to methods of treatment and diagnosis practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [X] Claims Nos.: 21,24
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-55 partially, 56 and 76 completely

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-55 partially; 56 and 76 completely

A polypeptide of undisclosed function corresponding to SEQ ID NO:1 and subject-matter relating thereto.

Inventions 2. to 20. Claims: 1-55 partially; 57-75 and 77-95 completely for inventions (SEQ ID NO) 2-20 respectively

Polypeptides of undisclosed function corresponding to SEQ ID NO:2-20 respectively and subject-matter relating thereto.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 21,24

Present claims 21 and 24 relate to compounds defined solely by reference to a desirable characteristic or property, namely the ability to agonise or antagonise some undefined characteristic of a polypeptide of the invention.

The claims cover all compounds having this characteristic or property, whereas the application provides no support within the meaning of Article 6 PCT nor disclosure within the meaning of Article 5 PCT for any such compound. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search of said claims is impossible. Independent of the above reasoning, said claims also lack clarity (Article 6 PCT) as an attempt is made to define the compounds referred to therein merely by reference to a result (agonism or antagonism) to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

nation on patent family members

International Application No
PCT/[redacted] 02/03709

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0028027 A	18-05-2000	AU 1457800 A	29-05-2000
		EP 1124958 A2	22-08-2001
		JP 2002529086 T	10-09-2002
		WO 0028027 A2	18-05-2000